AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

VOLUME LXX
1948

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX

THE ASYMPTOTIC ARCUS VARIATION OF SOLUTIONS OF REAL LINEAR DIFFERENTIAL EQUATIONS OF SECOND ORDER.*

By PHILIP HARTMAN and AUREL WINTNER.

1. Let t be a real independent variable which tends to $+\infty$, and let primes denote differentiations with respect to t. Consider the differential equation

(1) $$x'' + \omega^2 x = 0,$$

where $\omega = \omega(t)$ is a positive, continuous function defined for large t. Under suitable conditions on the behavior of $\omega = \omega(t)$ as $t \to \infty$, certain explicit asymptotic formulae are known for $x(t)$ and $x'(t)$, where $x = x(t)$ is an arbitrary solution of (1); cf. [1], [6], [7]. These asymptotic formulae imply, among other things, asymptotic formulae for the number of zeros of $x(t)$ on a large t-interval. In what follows, an asymptotic formula of this latter kind will be obtained under less restrictive conditions on $\omega(t)$; conditions which turn out to be the best possible of their kind.

(I) *Let $\omega = \omega(t)$, where $0 \leqq t < \infty$, be a positive, continuous function; $x = x(t)$ any real-valued non-trivial ($\not\equiv 0$) solution of* (1); *finally, $N(t)$ the number of zeros of $x(r)$ on the interval $0 \leqq r \leqq t$. Then the asymptotic rule*

(2) $$N(t) \sim \pi^{-1} \int_0^t \omega(r)\,dr \qquad (t \to \infty)$$

holds whenever $\omega(t)$ has a (continuous) derivative satisfying

(3) $$\omega' = o(\omega^2) \text{ as } t \to \infty.$$

Condition (3) is the best possible of its kind:

(I bis) *In* (I), *the relation* (2) *can fail to hold if the o in* (3) *is relaxed to O; in fact,*

(3 bis) $$\omega' = O(\omega^2) \text{ as } t \to \infty$$

and

(3*) $$\omega \to \infty \text{ as } t \to \infty$$

* Received May 20, 1947.

together are compatible with the failure of (2), *and not even the additional restriction that* ω *be monotone is of avail.*

As an illustration, choose

$$\omega(t) = t^{\alpha} \qquad (1 \leqq t < \infty),$$

where α is a (real) constant. Then (1) becomes a normal form of Bessel's equation (cf. [3]), condition (3) its satisfied if $\alpha > -1$ (but the restriction (3*), which is superfluous, is fulfilled only if $\alpha > 0$), and (2) gives

$$N(t) \sim \beta t^{\alpha+1}, \text{ where } \beta^{-1} = \pi(\alpha + 1) \qquad (\alpha > -1),$$

in agreement with the asymptotic formula of Bessel's functions.

2. Needless to say, if $N_1(t)$ and $N_2(t)$ are the functions $N(t)$ belonging to two different solutions, then, in view of Sturm's separation theorem, the absolute value of the difference $N_1(t) - N_2(t)$ cannot exceed unity and is, therefore, $O(1)$. On the other hand, in (2),

$$(4) \qquad \int_0^\infty \omega(t)\,dt = \infty \qquad (\omega > 0),$$

if (3), or for that matter just (3 bis), is assumed. In fact, (3 bis) means that the absolute value of the derivative of $1/\omega(t)$ is $O(1)$. Hence, $1/\omega(t) = O(t)$, and so $\omega(t) > \text{const.}/t$ holds for a positive constant as $t \rightarrow \infty$. This implies (4).

What will be used in the proof of (2) is somewhat less than the existence of the derivative $\omega'(t)$ and (3); namely, just

$$(5) \quad \epsilon(t) \equiv \underset{t<u<v<\infty}{\text{l.u.b.}} \, |\log \omega(v)/\omega(u)| \,/\, (1 + \int_u^v \omega(r)\,dr) \rightarrow 0 \text{ as } t \rightarrow \infty.$$

Since, if (3) is assumed,

$$\log \omega(v)/\omega(u) = \int_u^v \{\omega'(r)/\omega(r)\}\,dr = o(\int_u^v \omega(r)\,dr)$$

holds uniformly for $u < v < \infty$, as $u \rightarrow \infty$, it is clear that (3) is sufficient for (5).

With reference to the non-increasing, non-negative function $\epsilon(t)$ defined by (5), choose a positive, monotone function $X(t)$ satisfying

$$(6) \qquad X(t) \rightarrow \infty \text{ as } t \rightarrow \infty,$$

and

(7) $$X(t)\epsilon(t) \to 0 \text{ as } t \to \infty,$$

finally

(8) $$X(t)/\int_0^t \omega(r)\,dr \to 0 \text{ as } t \to \infty.$$

In view of (4), and since $\epsilon(t) \to 0$, such functions $X(t)$ exist. Define monotone sequence $t_0, t_1, t_2, \cdots$ by placing $t_0 = 0$ and

(9) $$\int_{t_{k-1}}^{t_k} \omega(r)\,dr = X(t_{k-1}) \qquad (k = 1, 2, \cdots).$$

According to (4) and (6), this defines t_k in terms of t_{k-1} in such a way th

(10) $$t_k \to \infty \text{ as } k \to \infty.$$

Since (5) implies that

$$|\log \omega(v)/\omega(u)| \leqq \epsilon(t_{k-1})(1 + \int_u^v \omega(r)\,dr) \text{ if } t_{k-1} \leqq u < v < \infty,$$

it follows from (9) that

$$|\log \omega(v)/\omega(u)| \leqq \epsilon(t_{k-1})(1 + X(t_{k-1})) \text{ if } t_{k-1} \leqq u < v \leqq t_k,$$

and so, from (5) and (7), that

$$\log \omega(v)/\omega(u) \to 0 \text{ as } k \to \infty, \text{ where } t_{k-1} \leqq u < v \leqq t_k.$$

This means that

(11) $$M_k/m_k \to 1 \text{ as } k \to \infty,$$

where

(12) $$M_k = \max \omega(t) \text{ and } m_k = \min \omega(t) \text{ for } t_{k-1} \leqq t \leqq t_k.$$

In addition,

(13) $$k = o(\int_0^{t_k} \omega(t)\,dt) \text{ as } k \to \infty.$$

This is clear from (6) and (9).

3. The truth of (2) will now be deduced by applying Sturm's con parison theorem to (1) on the successive intervals $t_0 \leqq t \leqq t_1$, $t_1 \leqq t \leqq t_2, \cdots$ By (12), the number of zeros of a solution of (1) on the interval $t_{k-1} \leqq t \leqq$ is not more than the number of zeros of a real-valued, non-trivial solution

$y'' + M_k^2 y = 0$. But this latter equation admits of the solution $y = \cos M_k t$. Consequently, the number of zeros of a solution $x = x(t)$ of (1) on $t_{k-1} \leqq t \leqq t_k$ is not greater than $\pi^{-1} M_k (t_k - t_{k-1}) + 1$. Hence, if $N(t)$ denotes the number of zeros defined before (2),

$$(14) \qquad N(t) \leqq \pi^{-1} \sum_{t_{k-1} \leqq t} M_k (t_k - t_{k-1}) + \sum_{t_{k-1} \leqq t} 1.$$

But it is clear from (13) that

$$(15) \qquad \sum_{t_{k-1} \leqq t} 1 = o\left(\int_0^t \omega(r)\, dr \right).$$

On the other hand, from (12) and (9),

$$M_k (t_k - t_{k-1}) = (M_k / m_k) m_k (t_k - t_{k-1}) \leqq (M_k / m_k) X(t_{k-1})$$

so that (8) and (10) imply

$$(16) \qquad M_k (t_k - t_{k-1}) = o\left(\int_0^{t_{k-1}} \omega(r)\, dr \right).$$

Thus, from (14), (15) and (16),

$$(17) \qquad \pi N(t) \leqq \sum_{t_k \leqq t} M_k (t_k - t_{k-1}) + o\left(\int_0^t \omega(r)\, dr \right).$$

Since, by the definition, (12), of m_k,

$$\int_0^t \omega(r)\, dr \geqq \sum_{t_k \leqq t} m_k (t_k - t_{k-1}),$$

it follows from (17) that

$$\pi N(t) / \int_0^t \omega(r)\, dr \leqq \sum_{t_k \leqq t} M_k (t_k - t_{k-1}) / \sum_{t_k \leqq t} m_k (t_k - t_{k-1}) + o(1),$$

as $k \rightarrow \infty$. Hence, by (11) and (10),

$$\limsup_{t \rightarrow \infty} \pi N(t) / \int_0^t \omega(r)\, dr \leqq 1.$$

It is clear that, if the "Sturmian majorant," $y'' + M_k^2 y = 0$, of (1) is replaced by the corresponding minorant, $y'' + m_k^2 y = 0$, the preceding deduction leads to

$$\liminf_{t\to\infty} \pi N(t) / \int_0^t \omega(r)\,dr \geqq 1,$$

and therefore to (2).

4. This completes the proof of the sufficiency of (3) for (2). What remains to be proved is the insufficiency of (3 bis) and (3*) together (even if $\omega(t)$ is to be monotone). The possible failure of (2) under the latter conditions will now be proved by an example. It results from the following construction:

Define a decreasing sequence of positive numbers $\epsilon_0, \epsilon_1, \cdots$ which tend to 0, and a sequence of positive integers $n_0, n_1, \cdots$, by placing $\epsilon_0 = 1$, $n_0 = 1$ and, if $\epsilon_1, \cdots, \epsilon_m$ and $n_1, \cdots, n_m$ have been defined, by choosing ϵ_{m+1} so as to satisfy the equalities $0 < \epsilon_{m+1} < \epsilon_m$ and

$$(18) \qquad \sum_{k=1}^{m} n_k/\epsilon_k + m < -\epsilon_m \log \epsilon_{m+1},$$

and then n_{m+1} so as to satisfy the inequality

$$(19) \qquad -\log \epsilon_{m+1} + \sum_{k=1}^{m} n_k/\epsilon_k + m < n_{m+1}.$$

In terms of these two sequences, define the coefficient function of (1) by placing

$$(20) \qquad \omega(t) = \begin{cases} 1/(T_k - t) & \text{if } T_k - \epsilon_k \leqq t < T_k - \epsilon_{k+1}, \\ 1/\epsilon_{k+1} & \text{if } T_k - \epsilon_{k+1} \leqq t < T_{k+1} - \epsilon_{k+1}, \end{cases}$$

where

$$(21) \qquad T_k = \sum_{j=0}^{k} n_j \qquad (k = 0, 1, \cdots).$$

This function $\omega(t)$, consisting of hyperbolic arcs and segments of constancy, is continuous; it has only right and left hand derivatives at a sequence of isolated points but, as far as the following considerations go, the curve $\omega = \omega(t)$ could be rounded off so as to have a continuous derivative throughout. For the sake of simplicity, this will be omitted; so that, at the corners, the derivative occurring in (3 bis) will be meant only as either of the unilateral derivatives.

In this sense, it is clear from (20) that $\omega'(t)$ is either $\omega^2(t)$ or 0; hence, (3 bis) is satisfied. Since $\omega'(t)$ is nowhere negative, $\omega(t)$ is monotone; hence, (3*) is satisfied, $\omega(t) = O(1)$ being prevented by the second line of (20), where $\epsilon_{k+1} \to 0$. Accordingly, what must be verified is that (2) fails to hold in the present case.

First, it is seen from (20) that, if t is on the interval

(22) $$T_k - \epsilon_{k+1} \leqq t < T_{k+1} - \epsilon_{k+1},$$

then (1) reduces to

$$x'' + x/\epsilon_{k+1}{}^2 = 0$$

and admits, therefore, of the solution

$$x(t) = \cos(t/\epsilon_{k+1}).$$

Since the number of zeros of this function on the interval (22) cannot differ by more than one from the product of $1/\pi\epsilon_{k+1}$ and the length of the interval (22), and since this length is

$$(T_{k+1} - \epsilon_{k+1}) - (T_k - \epsilon_{k+1}) = T_{k+1} - T_k = n_{k+1},$$

by (21), the contribution of the interval (22) to $N(t)$ is

(23) $$n_{k+1}/\pi\,\epsilon_{k+1} + O(1) \text{ as } k \to \infty$$

(in fact, $|O(1)| \leqq 1$).

In order to evaluate the contribution to $N(t)$ of the intervals complimentary to (22), that is, of the intervals

(24) $$T_k - \epsilon_k \leqq t < T_k - \epsilon_{k+1},$$

it is convenient to replace the independent variable t by the new independent variable

(25) $$s = s(t) = \int_0^t \omega(r)\,dr$$

(Liouville). Since $\omega(t)$ is positive, t and s are increasing functions of each other. Clearly, (25) transforms (1) into

(26) $$D(\omega Dx) + \omega x = 0,$$

where

(27) $$D = d/ds \qquad \text{(while } ' = d/dt\text{)}.$$

Since (25) and (27) imply that

(28) $$D(\omega Dx) = \omega D^2x + (D\omega)(Dx) = \omega D^2x + (\omega' Dx)/\omega,$$

(26) can be written as

(29) $$D^2x + (\omega'/\omega^2)Dx + x = 0.$$

In view of the first line of (20), this becomes

$$D^2x + Dx + x = 0$$

on that s-interval which, by virtue of (25), corresponds to the t-interval (24). The last differential equation admits of the solution

$$x(t(s)) = e^{-\frac{1}{2}s} \cos (3^{\frac{1}{2}}s/2).$$

Hence, in virtue of (25) and the first line of (20), the equation (1) admits of the solution

$$x(t) = (T_k - t)^{\frac{1}{2}} \cos \{(3^{\frac{1}{2}}/2) \log (T_k - t)\}$$

on the interval (24).

Consequently, the number of zeros on the interval (24) cannot differ by more than one from the product of $3^{\frac{1}{2}}/2\pi$ and

$$-\log (T_k - T_k + \epsilon_{k+1}) + \log (T_k - T_k + \epsilon_k) = \log (\epsilon_k/\epsilon_{k+1}).$$

Hence, the contribution of the interval (24) to $N(t)$ is

(30) $$(3^{\frac{1}{2}}/2\pi) \log (\epsilon_k/\epsilon_{k+1}) + O(1) \text{ as } k \to \infty$$

(in fact, $|O(1)| \leqq 1$).

For a fixed m, let k run from 0 to m in both (22) and (24). Then, since the contribution to $N(t)$ of the respective intervals is given by (23) and (30),

(31) $$\pi N(T_m - \epsilon_{m+1}) = -(3^{\frac{1}{2}}/2) \log \epsilon_{m+1} + \sum_{k=1}^{m} n_k/\epsilon_k + O(m)$$

and

(32) $$\pi N(T_{m+1} - \epsilon_{m+1}) = -(3^{\frac{1}{2}}/2) \log \epsilon_{m+1} + \sum_{k=1}^{m+1} n_k/\epsilon_k + O(m),$$

as $m \to \infty$. Clearly, (31) and (18) imply that

(33) $$\pi N(T_m - \epsilon_{m+1}) = -(3^{\frac{1}{2}}/2) \log \epsilon_{m+1} - O(\epsilon_m) \log \epsilon_{m+1},$$

while (32) and (19) imply that

(34) $$\pi N(T_{m+1} + \epsilon_{m+1}) = n_{m+1}/\epsilon_{m+1} + O(\epsilon_{m+1}) n_{m+1}/\epsilon_{m+1}.$$

On the other hand, the first and second lines of (20) show that the con-

tributions of the k-th of the intervals (22) and (24) to the integral which multiplies π^{-1} in (2) amount to

$$n_{k+1}/\epsilon_{k+1} \text{ and } \log(\epsilon_k/\epsilon_{k+1}),$$

respectively. Hence,

$$\int_0^{T_m-\epsilon_{m+1}} \omega(r)\,dr = -\log \epsilon_{m+1} + \sum_{k=1}^{m} n_k/\epsilon_k.$$

and

$$\int_0^{T_{m+1}-\epsilon_{m+1}} \omega(r)\,dr = -\log \epsilon_{m+1} + \sum_{k=1}^{m+1} n_k/\epsilon_k,$$

It follows, therefore, from (18) and (19), respectively, that

$$(35) \qquad \int_0^{T_m-\epsilon_{m+1}} \omega(r)\,dr = -\log \epsilon_{m+1} - O(\epsilon_m) \log \epsilon_{m+1}$$

and

$$(36) \qquad \int_0^{T_{m+1}-\epsilon_{m+1}} \omega(r)\,dr = n_{m+1}/\epsilon_{m+1} + O(\epsilon_{m+1}) n_{m+1}/\epsilon_{m+1}.$$

If (35) is compared with (33), and (36) with (34), it is seen that

$$\pi N(t) / \int_0^t \omega(r)\,dr$$

tends to $3^{\frac{1}{2}}/2$ or 1 according as t tends to ∞ through the values $t = T_m - \epsilon_{m+1}$ or $t = T_{m+1} - \epsilon_{m+1}$. Hence, (2) cannot hold.

5. If (1) is written as a system of differential equations of first order for $x = x(t)$ and $y = x'(t)$ and if

$$\phi(t) = \arg z(t), \text{ where } z(t) = x(t) + iy(t),$$

then $N(t)$, the number of zeros of x, is identical with the variation of the arcus ϕ (that is, with the number of completed circuits of the z-curve about the point $z = 0$), since

$$\phi = \operatorname{arc\,tan}\,(x'/x).$$

According to this interpretation (cf. [4]), the relation (2) refers to the asymptotic behavior of $x'(t)/x(t)$. Hence, the following theorem can be interpreted as a dual of (I):

(II) *The assertion of* (I) *remains true if*

(36) $$x'' - \omega^2 x = 0$$

and

(37) $$x'(t)/x(t) \sim \pm\, \omega(t) \qquad (t \rightarrow \infty)$$

are read instead of (1) *and* (2), *respectively.*

Again, (3) is the best condition of its kind:

(II bis) *In* (II), *the relation* (37) *can fail to hold if the assumption* (3) *of* (II) *is relaxed to* (3 bis); *in fact, not even the additional assumption* (3*) *and the monotony of* $\omega(t)$ *are of avail.*

In order to prove (II), define a new independent variable, s, by (25). Since $\omega(t)$ is positive, t and s are strictly increasing functions of each other, and

(38) $$s \rightarrow \infty \text{ as } t \rightarrow \infty,$$

by (4). Clearly, (25) transforms (1) into

$$D(\omega D x) - \omega x = 0,$$

where $D = d/ds$; cf. (27). In view of (28), the last equation can be written in the form

$$D^2 x + (\omega'/\omega^2) D x - x = 0.$$

This is a homogeneous linear differential equation

$$D^2 x + a(s) D x + b(s) x = 0,$$

in which $a(s)$ and $b(s)$ are (continuous) functions having finite limits $a(\infty) = 0$, $b(\infty) = -1$, the first of these relations being precisely the assumption (3). Since the limiting characteristic polynomial,

$$\lambda^2 + a(\infty)\lambda + b(\infty) = \lambda^2 - 1,$$

has the zeros $\lambda = \pm 1$, it follows that the logarithmic derivative of every non-trivial solution satisfies one of the limit relations

$$(Dx)/x \rightarrow \pm 1 \text{ as } s \rightarrow \infty$$

(Poincaré; Perron [5]). In view of (38) and (25), where $D = d/ds$, this is equivalent to (37), where $' = d/dt$.

This proves (II). In order to prove (II bis), it is sufficient to choose $\omega(t)$ in (36) to be the function (20), which was used in the proof of (I bis).

Actually, the verification of the failure of the alternative (27) in the case (20) can be avoided. In fact, it could be concluded from the results of [2] that *if* $\omega(t)$ *is a positive continuous function defined for large* t, *then* (5) *is necessary and sufficient in order that the alternative* (37) *hold for every non-trivial solution of* (36). Hence, (II bis) can be concluded from the fact that the function $\omega(t)$, defined by (20), is monotone, satisfies (3 bis) and (3*), but fails to satisfy (5).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

[1] M. Bôcher, "On regular singular points of linear differential equations of the second order whose coefficients are not necessarily analytic," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 40-52.

[2] P. Hartman, "Unrestricted solution fields of almost-separable differential equations," to appear *ibid.*

[3] F. Klein, *Vorlesungen über die hypergeometrische Funktion* (ed. O. Haupt), Berlin, 1933, p. 157.

[4] T. Levi-Civita, "Sur les équations linéaires à coefficients périodiques et sur le moyen mouvement du nœud lunaire," *Annales Scientifiques de l'École Normale Supérieure*, ser. 3, vol. 28 (1911), pp. 325-376.

[5] O. Perron, "Ueber lineare Differentialgleichungen, bei denen die unabhängige Variable reell ist," *Journal für die reine und angewandte Mathematik*, vol. 142 (1913), pp. 254-270; vol. 143 (1913), pp. 25-50.

[6] A. Wintner, "Asymptotic integrations of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272 (Appendix, pp. 270-272).

[7] ———, "On the normalization of characteristic differentials in continuous spectra," *Physical Review*, vol. 72 (1947), pp. 516-517.

ON THE CRITICAL POINTS OF FUNCTIONS POSSESSING CENTRAL SYMMETRY ON THE SPHERE.*

By J. L. Walsh.

The object of this note is briefly to consider rational and harmonic functions defined on a sphere (the Neumann sphere) and which in each pair of diametrically opposite points take the same value. We establish results concerning the location of the critical points, namely the zeros of the derivative of a rational function and the points where both first partial derivatives vanish for a harmonic function.

1. The symmetry required here is precisely the symmetry required in Klein's classical model of elliptic geometry. Theorem 1 is the analog of results already established for euclidean and hyperbolic geometries:

Theorem 1. *Let $f(z)$ be a rational function of z whose zeros and whose poles occur in pairs in diametrically opposite points of the sphere. Let P be an arbitrary point of the sphere, and let S denote the (open) hemisphere containing P whose pole is P. Suppose the great circle C through P separates all zeros of $f(z)$ in S but not on C from all poles of $f(z)$ in S not on C, where we assume $f(z)$ to have at least one zero or pole interior to S not on C. Then P is not a finite zero of $f'(z)$ unless it is a multiple zero of $f(z)$.*

Corresponding to an arbitrary rational function $f(z)$, either on the sphere or in the extended plane, we consider a fixed particle at each zero, repelling with a force equal to the inverse distance, and a fixed particle at each pole, attracting with a force equal to the inverse distance; multiple zeros and poles are represented by multiple particles. It then follows [1] that the finite critical points of $f(z)$ are precisely the positions of equilibrium in the field of force and the multiple zeros of $f(z)$.

We choose, as we may do with no loss of generality, the point P as the origin: $z = 0$ in the Gauss plane, and the image of S as S': $|z| < 1$. The image of C is a line C' through P.

If z_1 is a zero of $f(z)$, so is the point $-1/\bar{z}_1$, and the forces at P due

* Received May 21, 1947.

[1] Bôcher, *Proceedings of the American Academy of Arts and Sciences*, vol. 40 (1904), pp. 469-484; Walsh, *Transactions of the American Mathematical Society*, vol. 19 (1918); pp. 291-298.

to the corresponding particles are respectively $-1/\bar{z}_1$ and z_1; the sum of these forces is

$$(1)\qquad z_1 - (1/\bar{z}_1) = -z_1\left(\frac{1 - z_1\bar{z}_1}{z_1\bar{z}_1}\right),$$

so that, if z_1 lies interior to S', this force is directed from z_1 to P. If z_1 lies on the unit circle, this force is zero. Under the conditions of Theorem 1, each pair of zeros of $f(z)$ not on the unit circle has but one member z_1 interior to S'; the total corresponding force exerted at P due to the pair of particles is directed from z_1 toward P, hence if z_1 does not lie on C' this force has a non-vanishing component orthogonal to C'; each pair of poles of $f(z)$ not on the unit circle has but one member z_2 interior to S', and the total corresponding force exerted at P is directed from P toward z_2, hence if z_2 does not lie on C' this force has a non-vanishing component orthogonal to C' in the *same* sense as the component of the force due to each pair of zeros of $f(z)$ not on C'. Since $f(z)$ has at least one zero or pole interior to S' but not on C', the point P cannot be a position of equilibrium, hence cannot be a zero of $f'(z)$ unless it is a multiple zero of $f(z)$. Theorem 1 is established.

Theorem 1 is analogous to a result on rational functions in the euclidean plane [Bôcher and Walsh, *loc. cit.*], and to a result on rational functions in the hyperbolic plane.[2]

We note incidentally that if we modify the hypothesis of Theorem 1 so as to allow all zeros and poles of $f(z)$ to lie on the great circle one of whose poles is P, and if the zeros and poles of $f(z)$ are chosen on this great circle, then P is a position of equilibrium and critical point of $f(z)$. Likewise if the hypothesis of Theorem 1 is modified so as to permit all zeros and poles of $f(z)$ to lie on C', and if all zeros and poles of $f(z)$ do lie on C', then P may be a critical point of $f(z)$ even if not a multiple zero of $f(z)$; indeed P may be a point of symmetry for the zeros and poles of $f(z)$.

We proceed to a simple application of Theorem 1. By a *circular region* we mean a closed region of the plane or sphere bounded by a circle. Where confusion will not result, we use the same notation for a circular region as for its boundary. Let the function $f(z)$ possess the symmetry required in Theorem 1, let the zeros of $f(z)$ lie in two diametrically opposite circular regions C_1 and C_3 of the sphere, and the poles of $f(z)$ lie in two diametrically opposite circular regions C_2 and C_4 disjoint from C_1 and C_3. Construct the circular region S_k $(k = 1, 2, 3, 4)$ which is the locus of points (assumed not empty) P such that every point of the region C_k is at a spherical distance from P not greater than $\pi/2$. Thus the circular regions C_k and S_k have the

[2] Walsh, *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 462-470.

same poles, and the measures of their spherical radii are complementary. Denote by T the open region (assumed not empty) common to S_1 and S_2. If P is a point of T, the open hemisphere whose pole is P and which contains P contains the regions C_1 and C_2, and contains no point of C_3 or C_4. Taken together with suitably chosen arcs of C_1 and C_2, the great circles Γ_1 and Γ_2 tangent to C_1 and C_2 and separating those circles separate T into three regions R_1, R_2, R, of which R_1 and R_2 may be empty. The region R consists of all points P in T which lie on great circles which separate C_1 and C_2, and R is open. The region R_1 closed with respect to T is disjoint from R, and consists of all points of T not in R which lie on great circles cutting both C_1 and C_2, but on the maximal arcs of those circles bounded by points of C_1 and of the boundary of T, arcs in T containing no points of C_2; the region R_2 is similarly defined by permuting subscripts. The regions R_1 and R_2 may contain the whole or parts of the regions C_1 and C_2 respectively. Both R_1 and R_2 are convex with respect to great circles. It is an immediate consequence of Theorem 1 that *all finite critical points of $f(z)$ in T lie in R_1 and R_2; no finite critical point of $f(z)$ lies in R.* We have constructed the geometric configurations C_1, C_2, R_1, R_2, R, T for the purpose of simplicity of exposition; under suitable conditions corresponding regions R_1, R_2, R, T may be used without the introduction of circular regions C_k.

Another result is also not difficult to establish:

THEOREM 2. *Let $f(z)$ be a rational function of z whose zeros and whose poles occur in pairs in diametrically opposite points of the sphere. Let P be an arbitrary point of the sphere, and let S (considered closed) denote the hemisphere containing P whose pole is P. Let all zeros [or poles] of $f(z)$ in S lie exterior to the circular region C_1 containing P and bounded by a small circle in S of the sphere whose pole is P. Let all poles [or zeros] of $f(z)$ in S lie in the circular region C_2 interior to C_1 but not containing P. Then P is not a finite critical point of $f(z)$.*

We prove Theorem 2 by interpreting the configuration in the plane instead of on the sphere with the image of P (also denoted by P) in $z = 0$, the image of S as $S' : |z| \leqq 1$, the image of C_1 as $C'_1 : |z| \leqq \rho_1$, and the image of C_2 as C'_2. For definiteness suppose the zeros of $f(z)$ in S to lie exterior to C_1.

Let $2n$ denote the degree of $f(z)$. For each zero z_1 of $f(z)$ interior to S' it follows from (1) that the force at P for the corresponding pair of zeros of $f(z)$ is in magnitude less than

$$\frac{1}{\rho_1} - \rho_1,$$

so that the total force at P due to particles at all the zeros of $f(z)$ is less than $n(1-\rho_1^2)/\rho_1$.

Construct an auxiliary circle C_0 containing in its interior the interior of C'_2, so that C_0 passes through P and lies interior to C'_1; we assume for definiteness the center of C_0 to be the point $\rho/2$, on the positive half of the axis of reals. If a pole (r, θ) of $f(z)$ lies in or on C_0 $(r \neq 0)$, the force exerted at P by the corresponding pair of particles is in magnitude $(1/r) - r$, and the component of this force in the positive horizontal direction is $(1-r^2)(\cos\theta)/r$. But since the pole lies in or on C_0 we have

$$1 \geqq \cos\theta \geqq \frac{r}{\rho} \geqq \frac{r(1-\rho^2)}{\rho(1-r^2)}, \quad \frac{(1-r^2)\cos\theta}{r} \geqq \frac{1}{\rho} - \rho;$$

this inequality merely expresses the fact that for each pair of poles, of which one representative lies in or on C_0 (but not at P), the corresponding force exerted at P has a horizontal component in the positive sense not less than it would be if the representative coincided with the point $(\rho, 0)$. Thus under the conditions of Theorem 2 the total force at P due to the particles at the $2n$ poles has a horizontal component in the positive sense which is greater than the total force at P due to the particles situated at the $2n$ zeros of $f(z)$. Thus P cannot be a position of equilibrium nor a multiple zero of $f(z)$, so that P is not a critical point of $f(z)$.

We mention an immediate application of Theorem 2. Let $f(z)$ have the symmetry required in Theorem 2, and let all zeros [or poles] of $f(z)$ lie in diametrically opposite circular regions C_1 and C_3 of the sphere. Let no pole [or zero] of $f(z)$ lie in the circular region C_2 which has the same poles as C_1 and which contains C_1. Denote by ρ_1 and ρ_2 $(< \pi/2)$ the spherical radii of C_1 and C_2, and suppose $\rho_2 > 3\rho_1$. Then *the (open) zone bounded by C_1 and the circle C in C_2 having the same poles as C_1 and spherical radius $(\rho_2 - \rho_1)/2$ contains no finite critical points of $f(z)$.* If P is an arbitrary point of this zone, whose spherical distance from the pole of the region C_1 is denoted by ρ, the circular region containing P whose pole is P and spherical radius $\rho + \rho_1$ contains C_1 but contains no point of the circle C_2; the hemisphere containing P whose pole is P contains no point of C_3; the conclusion follows from Theorem 2.

In the proof of Theorem 2 we notice that if Γ is the great circle through P orthogonal to the great circle through P orthogonal to C_2, then the zeros of $f(z)$ in S which are on Γ or separated from C_2 by Γ need not lie exterior to C_1; the proof already given is valid if we notice that pairs of zeros represented by points in S separated from C_2 by Γ yield a non-negative component at P in the direction and sense from P toward the center of C_2. Thus we have the

COROLLARY. *Let $f(z)$ have the symmetry required in Theorem 2, let P be an arbitrary point of the sphere, and let S be the closed hemisphere containing P whose pole is P. Let each zero [or pole] of $f(z)$ in S lie either exterior to the circular region C_1 containing P and bounded by a small circle of the sphere in S whose pole is P, or lie not at P but on a great circle Γ through P, or be separated by Γ from a circular region C_2 in S not containing P whose boundary is orthogonal to the great circle through P orthogonal to Γ. Let all poles [or zeros] of $f(z)$ in S lie in C_2. Then P is not a finite critical point of $f(z)$.*

As a simple application of Theorems 1 and 2 we establish

THEOREM 3. *Let $f(z)$ be a rational function of z whose zeros and whose poles occur in pairs in diametrically opposite points of the sphere. Let all the zeros lie in diametrically opposite circular regions C_1 and C_3 and all the poles lie in diametrically opposite circular regions C_2 and C_4, where all the regions C_k are mutually disjoint. Denote by Z_1 and Z_2 the (closed) zones of the sphere which are the loci of the great circles whose poles lie in C_1 and C_3 and in C_2 and C_4 respectively. Suppose that for the circles C_1 and C_2 (and likewise C_2 and C_3), the (spherical) length of the common tangent T (chosen as the shorter arc of a great circle separating the regions C_1 and C_2) is not less than the sum of the (spherical) diameters of those circles.*[3] *Then all finite zeros of $f'(z)$ lie in the regions C_k and in Z_1 and Z_2.*

Let a point P of the sphere not in a region C_k or a zone Z_k be a finite zero of $f'(z)$; we shall reach a contradiction. The hemisphere S whose pole is P and which contains P must contain in its interior the whole of one of the regions C_1 and C_3, and no point of the other of those regions, for P does not lie in Z_1; similarly for the regions C_2 and C_4. Let us suppose S to contain C_1 and C_2.

No great circle through P can separate C_1 and C_2, by Theorem 1. For definiteness suppose the pole of C_1 in S to be nearer P than that of C_2 in S. There exists a great circle C through P which cuts C_1 and C_2 at supplementary angles, those circles being oriented in the same sense on the sphere; thus C may be defined as the great circle through P and through the intersection of the common tangents T to C_1 and C_2. Then C cuts C_1 and C_2 in such a way that the points P, A_1, B_1, A_2, B_2 lie in that order on an arc of C in S, where A_k and B_k lie on C_k; the points A_k and B_k do not coincide unless C is tangent

[3] This condition is merely a convenient one for use in the proof; it may be replaced by other conditions less restrictive.

to C_k at A_k. The spherical distance A_1B_2 is not less than the length of T, which by hypothesis is not less than the sum of the diameters of C_1 and C_2. The circle Γ whose pole is P and whose radius is PB_2 less the spherical measure of the diameter of C_2 cannot cut C_2. Also, Γ cuts C between A_1 and B_2 at a point whose distance from A_1 is not less than the diameter of C_1, so that Γ cannot cut C_1, and C_1 is separated by Γ from C_2. It follows from Theorem 2 that P is not a critical point of $f(z)$; this contradiction completes the proof of Theorem 3.

2. Theorems 1 and 2 are both results which, referring to the zeros and poles of $f(z)$ located in a hemisphere S which contains its pole P, assert that if the zeros of $f(z)$ lie in a suitable circular region C_1 and if the poles of $f(z)$ lie in a circular region C_2 which has no point in common with C_1 and satisfies auxiliary conditions, then P is not a critical point of $f(z)$. The question suggests itself whether some auxiliary conditions are here necessary, a question which we answer in the affirmative:

THEOREM 4. *There exists a rational function $f(z)$ of degree four whose zeros and whose poles occur in pairs in diametrically opposite points of the sphere; there exists a point P which is a finite critical point of $f(z)$; there exist two diametrically opposite circular regions C_1 and C_2 each of which contains precisely two zeros of $f(z)$ and which do not contain P nor intersect the great circle Γ one of whose poles is P; the poles of $f(z)$ are double and lie exterior to C_1 and C_2.*

Theorem 4 is trivial if we omit the requirement that C_1 and C_2 shall not intersect Γ, for if P is arbitrary and all zeros and poles of $f(z)$ lie on Γ, then P is a position of equilibrium and hence a critical point of $f(z)$.

Let P be given, together with the arbitrary diametrically opposite circular regions C_1 and C_2 which neither contain P nor intersect Γ. We fasten our attention on the representatives in C_1 of the pairs of zeros of $f(z)$. Given two zeros z_1 and z_2 of $f(z)$ in C_1, the corresponding two pairs of zeros of $f(z)$ can be replaced by a single pair of double zeros (possessing the required symmetry) without altering the corresponding force exerted at P. If P is the point $z = 0$ and Γ is the unit circle, the equation to be solved for one of these double zeros z_0 is

$$(z_1 - 1/\bar{z}_1) + (z_2 - 1/\bar{z}_2) = 2(z_0 - 1/\bar{z}_0). \tag{2}$$

It will be noticed that equation (2) with the first member not zero defines z_0 in $|z_0| < 1$ uniquely; we merely write

$$z_0\left(\frac{1-z_0\bar{z}_0}{z_0\bar{z}_0}\right)=A\neq 0, \qquad \arg z_0=\arg A,$$

$$z_0\bar{z}_0\left(\frac{1-z_0\bar{z}_0}{z_0\bar{z}_0}\right)^2=A\bar{A};$$

this last equation defines $|z_0|$ uniquely, subject to the condition $|z_0|<1$. If for suitable choice of C_1, z_1, and z_2, the point z_0 cannot be chosen in the region C_1, we need merely choose the double poles of $f(z)$ in z_0 and in $-1/\bar{z}_0$ in order to establish Theorem 4.

The entire question, now reduced to the location of the point z_0, can be further transformed. To every point z_1 of the circular region C_1 we make correspond the point $w_1=z_1-(1/\bar{z}_1)$; if the locus of the points w_1 is convex, then to every pair of points z_1 and z_2 in C_1 corresponds by (2) a point z_0 also in C_1 equivalent to z_1 and z_2 in the sense that a double particle at z_0 (together with its mate at $-1/\bar{z}_0$) exerts the same force at P as do the pairs corresponding to z_1 and z_2. On the other hand, if the locus of the points w_1 is not convex, then for suitably chosen z_1 and z_2 in C_1 the pair z_0 and $-1/\bar{z}_0$ defined by (2) has no representative in C_1, and the double poles of $f(z)$ in Theorem 4 may be chosen in the points z_0 and $-1/\bar{z}_0$ *exterior to* C_1, and Theorem 4 is established. We proceed, then, to study the convexity of the locus of w_1. Here it is a slight convenience algebraically to choose C_1 as exterior to $|z|=1$; this is a choice merely of studying one of the two members of a symmetric pair rather than the other.

If the point z is exterior to $|z|=1$, the point w defined by the equation $w=z-(1/\bar{z})=z(z\bar{z}-1)/z\bar{z}$ lies on the half line from the origin through z, and (it is sufficient here to investigate real z) $|w|$ increases as $|z|$ increases. Let C_1 be the circle $|z-a|=r$ with $a>1+r$, since C_1 is exterior to the unit circle. It is no loss of generality to choose a real, which we do, setting $z=a+re^{i\theta}$, $w=u+iv$. Straightforward algebraic computation of $du/d\phi$ and $dv/d\phi$ then yields

$$\frac{dv}{du}=\frac{-(a^2+r^2)\cos\phi-2ar+\cos\phi(a^2+r^2+2ar\cos\phi)^2}{-\sin\phi[(a^2-r^2)+(a^2+r^2+2ar\cos\phi)^2]}.$$

This denominator vanishes only when $\sin\phi=0$, for we have $a>r$. Further computation shows that the algebraic sign of $d(dv/du)/d\phi$ *is* the same as the algebraic sign of the function

$$\begin{aligned}F(\phi)=(a^2+r^2+2ar\cos\phi)^3-2r(r+a\cos\phi)(a^2+r^2+2ar\cos\phi)\\+8a^2r\sin^2\phi(r+a\cos\phi)-(a^2-r^2);\end{aligned}$$

in writing this equation we have suppressed a factor $a^2+r^2+2ar\cos\phi$,

which is equal to $z\bar{z}$ and positive. It may be verified that $F(0)$ and $F(\pi)$ are both positive, and that when r is small in comparison with a the function $F(\phi)$ is positive for all values of ϕ.[4]

We now choose specifically $r = 1$, $\cos\phi = -2/a$, and set $a = 2 + \epsilon$, $\epsilon > 0$. Then we have $F(\phi) = -16\epsilon + \cdots$, where only powers of ϵ higher than the first are omitted. Consequently for suitably chosen ϵ we have $F(\phi) < 0$, the locus of w is not convex, and Theorem 4 is established.

3. Theorems 1, 2, and 3 have application to the study of harmonic functions defined on the surface of the sphere and possessing there suitable symmetry. The analog of Theorems 1 and 2 is

THEOREM 5. *Let R be a region of the sphere bounded by a finite number of mutually disjoint Jordan curves, and let R possess central symmetry. Let the function $U(x, y)$ be harmonic in R, continuous in the corresponding closed region, equal to zero on a symmetric set J_0 and to unity on the remaining set J_1 of the curves bounding R. Let P be an arbitrary point of the sphere, and let S denote the closed hemisphere containing P whose pole is P.*

Suppose that the great circle C through P separates all points of J_0 in S from all points of J_1 in S. Then P is not a critical point of $U(x, y)$.

Suppose that all points of J_0 [or of J_1] in S lie exterior to the circular region C_1 containing P and bounded by a small circle in S of the sphere whose pole is P. Suppose that all points of J_1 [or of J_0] in S lie in the circular region C_2 interior to C_1 but not containing P. Then P is not a critical point of $U(x, y)$.

The function $U(x, y)$ is the harmonic measure of J_1 in the point (x, y) with respect to the region R, and the function $1 - U(x, y)$, which has the

[4] Indeed, $F(\phi)$ is positive whenever $r + a\cos\phi$ is positive. Thus for suitably restricted circles C_1 and also for suitably chosen other regions R whose boundaries may contain arcs of circles C_1, *the locus of w is convex.* Under such conditions let us say that *R has Property α*; this property here depends on a particular point P. If the rational function $f(z)$ has the requisite symmetry, if P is an arbitrary point, if S denotes the closed hemisphere containing P whose pole is P, and if disjoint closed regions R_1 and R_2 interior to S not containing P and having property α contain respectively all zeros and all poles of $f(z)$ in S, then P is not a critical point of $f(z)$; compare Theorem 3. By way of proof it is sufficient to note that in the field of force, without altering the total force exerted at P, all positive particles in S may be concentrated at a single point which lies in R_1 and all negative particles in S may be concentrated at a single point which lies in R_2, without altering the symmetry. Thus the total force at P cannot vanish, and P cannot be a critical point of $f(z)$.

same critical points as $U(x,y)$, is the harmonic measure of J_0 in the point (x,y) with respect to R.

We represent $U(x,y)$ by integrals taken over the sets of curves J_0 and J_1 if these curves are analytic, and otherwise taken over neighboring sets of analytic level curves of $U(x,y)$, curves J'_0 and J'_1 in R found from J_0 and J_1 by slight deformations:[5]

$$U(x,y)-\kappa=-(1/2\pi)\int_{J'_0}\log r(\partial U/\partial\nu)\,ds-(1/2\pi)\int_{J'_1}\log r(\partial U/\partial\nu)\,ds,\quad (x,y) \text{ in } R;$$

here ν indicates exterior normal, and κ is a suitable constant. If we introduce the new variable σ by setting

$$d\sigma=-(\partial U/\partial\nu)\,ds \text{ on } J'_0,\quad d\sigma=(\partial U/\partial\nu)\,ds \text{ on } J'_1,\quad \tau=\int_{J'_0}d\sigma=\int_{J'_1}d\sigma,$$

we may write

$$\begin{aligned} U(x,y)-\kappa &= \frac{1}{2\pi}\int_0^\tau \log r\,d\sigma-\frac{1}{2\pi}\int_\tau^\tau \log r\,d\sigma \\ &= \frac{\tau}{2\pi}\lim_{n\to\infty}\frac{\log r_{n1}+\log r_{n2}+\cdots+\log r_{nn}}{n} \\ &\quad -\frac{\tau}{2\pi}\lim_{n\to\infty}\frac{\lim \rho_{n1}+\log\rho_{n2}+\cdots+\log\rho_{nn}}{n} \end{aligned}$$

where $r_{nk}=|z-\alpha_{nk}|$, $\rho_{nk}=|z-\beta_{nk}|$, and α_{nk} and β_{nk} are suitably chosen on J'_0 and J'_1 respectively. Convergence is uniform on any closed preassigned set interior to R.

By virtue of the symmetry of R and $U(x,y)$, the sets of curves J'_0 and J'_1 may also be chosen to be symmetric in the center of the sphere, and likewise the sets of points α_{nk} and β_{nk}, where n is taken as even. In the neighborhood of an arbitrary point of R, a constant multiple of the function $U(x,y)-\kappa$ is thus the uniform limit of the logarithm of a sequence of rational functions $R_n(z)$ whose zeros and poles lie outside of that neighborhood; indeed the functions $R_n(z)$ possess the symmetry demanded of $U(x,y)$, and their zeros and poles can be chosen to lie respectively in the regions required in Theorem 5 for J_0 and J_1. Theorems 1 and 2 apply to the functions $R_n(z)$. If a point P satisfies the conditions of Theorem 5, so do all points in a suitable neighborhood of P; no point of such a neighborhood is a critical point of an $R_n(z)$. Each critical point of $U(x,y)$ in R is a limit of critical points of the $R_n(z)$, so that Theorem 5 follows.

---

[5] Walsh, *Interpolation and Approximation* (New York, 1935), § 8.7.

The reader will have no difficulty in formulating and proving the analog of Theorem 3.

4. The relation between Theorem 1 and its analog in hyperbolic geometry is closer than mere analogy. Theorem 1 deals with the point $P: z = 0$ as a possible critical point of the rational function $f(z)$, whose zeros are the points α_k and $-1/\alpha_k$ $(k = 1, 2, \cdots, n)$, and whose poles are the points β_k and $-1/\bar{\beta}_k$ $(k = 1, 2, \cdots, n)$; we choose this notation so that we have $|\alpha_k| \leqq 1$, $|\beta_k| \leqq 1$. So far as concerns the force exerted at P, the field of force corresponding to $f(z)$ is equivalent to the field of force corresponding to the rational function $F(z)$ whose zeros are the points α_k and $1/\bar{\beta}_k$, and whose poles are the points β_k and $1/\bar{\alpha}_k$ $(k = 1, 2, \cdots, n)$; we omit from this enumeration points α_k, β_k, $1/\bar{\alpha}_k$, and $1/\bar{\beta}_k$ of modulus unity. The zeros and poles of $F(z)$ interior to the unit circle C are precisely those of $f(z)$ interior to C; the zeros and poles of $F(z)$ exterior to C are precisely the negatives of the poles and zeros respectively of $f(z)$ exterior to C. It will be noted that the zeros of $F(z)$ are the inverses of the poles of $F(z)$ with respect to C; the function $F(z)$ is of constant modulus on C. The point P is a critical point of $f(z)$, whether a multiple zero or a position of equilibrium in the field of force, when and only when P is a critical point of $F(z)$. It is thus a consequence of Theorem 1 that *if a line L separates the zeros of $F(z)$ interior to C: $|z| = 1$ not on L from the poles of $F(z)$ interior to C not on L, and if C contains at least one zero or pole not on L, then P is not a critical point of $F(z)$ unless P is a multiple zero of $F(z)$.* The property possessed by $F(z)$, that its zeros are inverse with respect to its poles in a circle, is characteristic of all rational functions having constant modulus on a circle, and is invariant under linear transformation. We thus have from Theorem 1 by proceeding from a given $F(z)$ to the function $f(z)$, a result on a rational function $F(z)$ of constant modulus on an arbitrary circle, a result involving the images of lines L through P and the analog (*loc. cit.*) of Theorem 1 for hyperbolic geometry:

If the rational function $F(z)$ has its poles inverse to its zeros in a circle C, if P is an arbitrary point interior to C, and if a non-euclidean line L through P separates all the zeros of $F(z)$ interior to C not on L from all the poles of $F(z)$ interior to C not on L, where we suppose at least one such zero or pole to exist, then P is not a critical point of $F(z)$ unless P is a multiple zero of $F(z)$.

At first sight it might appear as if this last result were established only for a rational function $F(z)$ of special type, having the same number of zeros

as poles interior to C. However, in proceeding from a given $F(z)$ to the corresponding $f(z)$, we may provide the latter artificially with any number of zeros or poles situated on C itself and possessing the required symmetry. With this interpretation, these two analogous theorems of the elliptic and hyperbolic geometries respectively can be considered equivalent, in the sense that either can be trivially proved from the other.

The result on hyperbolic geometry is of some interest in the theory of functions, for by a suitable conformal map it applies to the study of a function meromorphic in an arbitrary simply connected region, of constant modulus on the boundary.

Results of the present note other than Theorem 1 can also be applied by the reader in the study of hyperbolic geometry. We state merely the analog (the equivalent) of Theorem 2: *Let R be a simply connected region provided with a non-euclidean hyperbolic geometry by means of a conformal map onto a circle. Let P be a point of R, let R_0 be the closed subregion of R containing P bounded by a (non-euclidean) circle whose (non-euclidean) center is P, and let R_1 be a closed subregion of R_0 not containing P bounded by a (non-euclidean) circle. If $f(z)$ is meromorphic in R and of constant modulus on the boundary of R, if the zeros [or poles] of $f(z)$ in R are n in number and lie exterior to R_0, and if the poles [or zeros] of $f(z)$ in R are not fewer than n in number and lie interior to R_1, then P is not a finite critical point of $f(z)$.*

HARVARD UNIVERSITY.

ON THE LOCATION OF CONTINUOUS SPECTRA.*

By Aurel Wintner.

The following considerations collect a few rules which in various cases lead to the possible position of the essential portion of the *spectrum* of a differential equation

$$y'' + (\lambda + f(x))y = 0, \qquad 0 \leqq x < \infty, \tag{1}$$

in a direct manner, namely, in terms which do not postulate the delicate knowledge needed for the determination of *spectral forms* (*densities*). The latter, which concern "normalizations" (Hellinger), depend on *sharp* estimates of asymptotic behavior; estimates which are not at all, or not easily, available in many cases in which the location of the set formed by the continuous spectrum and the cluster points of the point spectrum (that is, by the set representing the "essential" portion of the spectrum) can be determined *a priori*. The method will consist of appropriate adaptations of that indicated in [5] for the particular case in which the $f(x)$ in (1) is a lattice potential (a case in which the range, $0 \leqq x < \infty$, of (1) becomes replaced by $-\infty < x < \infty$).

The possibility of such a direct approach is due to the fact that the spectrum itself can be defined without an involvement of the Hilbert-Hellinger theory of spectral forms (the points of increase of the latter define the spectrum itself). Needless to say, what has a spectrum, or a spectral form, is not (1) itself, where $f(x)$ is any given real-valued continuous function, but is represented by (1) and the boundary condition together. On the other hand, the boundary condition is two-fold: for $x = \infty$, it is assigned by the (L^2)-requirement of Hilbert's space, whereas for $x = 0$ it is any homogeneous, linear assignment,

$$y(0)\cos\theta + y'(0)\sin\theta = 0; \tag{2}$$

for instance, $y(0) = 0$ or $y'(0) = 0$, where $\theta = 0$ and $\theta = \frac{1}{2}\pi$, respectively. Accordingly, the set of λ-values representing the spectrum depends on the θ occurring in (2). However, as pointed out by Weyl ([3], p. 251), what

* Received May 1, 1947.

was above called the essential portion of the spectrum is independent of θ. Consequently, it is possible to speak of the essential portion of the spectrum of (1) itself.

The direct definition of the spectrum of (1) (belonging to a fixed θ), that is, the definition not involving Green functions and spectral forms, is based on the totality of the inhomogeneous equations belonging to ((1),

$$y'' + (\lambda + f(x))y = g(x), \tag{3}$$

where $f(x)$ is a fixed, and $g(x)$ an unspecified, continuous function on the half-line $0 \leqq x < \infty$. In fact, with reference to a fixed θ in (2), the spectrum of (1) can be defined to be the set of those real λ-values corresponding to which there exists *some* function $g(x)$ of class (L^2) having the property that no solution $y(x) = y_\lambda(x)$ of (3) satisfying (2) is a function of class (L^2) (since f, g, y are continuous throughout, an (L^2)-condition affects, of course, the behavior at $x = \infty$ only). This definition of the *spectrum* (cf. [3], p. 251) agrees with that based on spectral forms (Hilbert, Hellinger; Toeplitz). Correspondingly, the *point spectrum* of (1) (and of (2), where θ is fixed) must be defined as the set of those real λ-values corresponding to which the homogeneous differential equation (1) has a solution $y(x) \not\equiv 0$ of class (L^2). It should be noted that $f(x)$ is not required to be of class (L^2) in either of these definitions.

It is known that, for every fixed θ, the spectrum is a closed set; that the point spectrum is contained in the corresponding spectrum; finally, that the set of the cluster points of the spectrum, being identical with what above was called the essential portion of the spectrum, is independent of θ.

Actually, the above definitions of "point spectrum" and "spectrum" are meaningful only if $f(x)$ is not of the *Grenzkreis* type, that is, only if not every solution of (2) is of class L^2 (for some λ, but then, by necessity, for every λ; cf. [3], p. 238). It will, however, be part of the first of the assertions to be proved, that the assumption to be imposed on $f(x)$ prevents the occurrence of this degenerate case.

The first of the criteria in question is one based on *stability* (*in phase space*), since it can be formulated as follows:

(I) *If $f(x)$, where $0 \leqq x < \infty$, is a continuous function satisfying*

$$-\infty \leqq \limsup_{x \to \infty} f(x) < \infty,$$

and if there exists a parameter value λ_0 corresponding to which both

$$y(x) = O(1) \text{ and } y'(x) = O(1) \qquad (x \to \infty) \tag{4}$$

hold for every solution (that is, for a pair of linearly independent solutions) of the case $\lambda = \lambda_0$ of (1), *then* (1) *is of the* Grenzpunkt *type, and λ_0 is a cluster point (hence, a point) of the spectrum of* (1).

In other words, λ_0 is either in the continuous spectrum or in the derivative of the point spectrum (perhaps in both).

Here and in the sequel, all data and solutions are understood to be real-valued.

It should be noted that the stability assumption of (I) concerns the phase space, (y, y'), rather than just the configuration space, (y), of

$$y'' + (\lambda_0 + f(x))y = 0. \tag{5}$$

The purpose opposite to that of (I) is served by the following criterion:

(II) *Let $f(x)$ be continuous on the half-line $0 \leqq x < \infty$ and let there exist a parameter value λ_0 corresponding to which the differential equation* (5) *possesses two linearly independent solutions, $y = y_1(x)$ and $y = y_2(x)$, having the following property: For every function $g(x)$ of class (L^2) on the half-line,*

$$g^*(x) = y_1(x) \int_x^\infty y_2(t) g(t)\,dt + y_2(x) \int_0^x y_1(t) g(t)\,dt \tag{6}$$

is a function of class (L^2) on the half-line (which implies, of course, that $y_2(x)$ is of class (L^2) there). Then, if (1) *is of the* Grenzpunkt *type, λ_0 cannot be a cluster point of the spectrum of* (1).

In other words, λ_0 is either not in the spectrum of (1) (for any choice of θ in (2)) or, if it is (for some θ), then it is in the point spectrum (for the same θ); in which case λ_0 is an isolated point of the spectrum (for the same θ).

Clearly, (6) can be written in the form

$$g^*(x) = \int_0^\infty K(x, t) g(t)\,dt, \tag{6 bis}$$

where

$$K(x, t) = K(t, x), \qquad 0 \leqq x < \infty;$$

and what (II) requires is that this symmetric kernel be bounded (in the sense of Hilbert's theory of bounded linear transformations).

As an illustration (which covers various cases investigated in the

literature, as well as more elaborate cases, not available to the usual treatments, which are based on explicit or asymptotic formulae), suppose that (1) has the following property: For some λ_0, the differential equation (5) has two linearly independent solutions corresponding to which there exist five positive constants satisfying

$$Ae^{cx} < y_1(x) < Be^{cx} \qquad \text{and} \qquad Ce^{-cx} < y_2(x) < De^{-cx},$$

as $x \to \infty$. Then (1) is of the *Grenzpunkt* type, and λ_0 cannot be a cluster point of the spectrum of (1).

In order to conclude this from (II), it is sufficient to observe that, by virtue of the last formula line, the expression on the right of (6) becomes majorized by an integral (6 bis) in which $K(x, t)$ is represented by a constant multiple of $\exp(-c\,|\,x - t\,|)$ (and g by $|\,g\,|$). In fact, it then follows that it is sufficient to ascertain the boundedness of the "Laurent kernel" $K(x, t) = k(|\,x - t\,|)$ in the case $k(s) = \exp(-c\,|\,s\,|)$, where $c > 0$. But the boundedness of this particular kernel (Weyl-Picard) is the simplest instance of the integral variant of the general "Laurent" (or, rather, "Fourier") criterion of Toeplitz; cf. [1], pp. 354-356.

Proof of (I). Suppose that the first assertion of (I) is false. Then (1) belongs to the *Grenzkreis* type. This means that (1) has two linearly independent solutions of class (L^2) for some and/or every λ; hence, for the particular $\lambda = \lambda_0$ occurring in (5). It follows, therefore, from the second of the two assumptions (4), that (5) has two linearly independent solutions, say $y = y_1(x)$ and $y = y_2(x)$, having a Wronskian which is the sum of two terms each of which is of the form $O(1)y_i(x)$, where both functions $y_i(x)$ are of class (L^2). Consequently, the Wronskian itself is of class (L^2). This contains, however, a contradiction. For, on the one hand, the Wronskian of any two linearly independent solutions of (5) is a non-vanishing constant and, on the other hand, a non-vanishing constant is not a function of class (L^2), the domain of integration being the half-line $0 \leqq x < \infty$.

This contradiction proves the first of the assertions of (I). It may be noted that neither the second of the two assumptions of (4) nor the limsup-assumption of (I) was used thus far.

In order to prove the second assertion of (I), suppose that this assertion is false. Then λ_0 is not a cluster point of the spectrum of (1). Hence, λ_0 is either *not in the spectrum for any* choice of θ in (2) or else λ_0 is *in the point spectrum for some* choice of θ in (2) (and λ_0 represents, in the second case, an isolated point of the spectrum belonging to the θ in question). These two possibilities will have to be ruled out separately.

The second of these possibilities can be disposed of as follows: Its assumption implies that (5) has a solution $y = y_0(x) \not\equiv 0$ of class (L^2). It follows, therefore, from the limsup-assumption of (I) and from Theorem (i) in [4], that the derivative $y'_0(x)$ is of class (L^2). On the other hand, since $y_0(x) \not\equiv 0$, it is possible to choose a solution $y = y^0(x)$ of (5) which is linearly independent of $y = y_0(x)$. Then the Wronskian of $y_0(x)$ and $y^0(x)$ is a non-vanishing constant. Hence, if the O-assumptions (4) are applied to $y^0(x)$ and its derivative, then, since $y_0(x)$ and its derivative are of class (L^2), there results the same contradiction as above.

What remains to be ruled out is the first possibility, that in which λ_0 is not in the spectrum for any choice of θ in (2). In view of that definition of the spectrum (belonging to a fixed θ) which is based on (3), this possibility can be characterized as follows: For every continuous $g(x)$ of class (L^2) and for every choice of θ in (2), the case $\lambda = \lambda_0$ of (3) has a solution of class (L^2) satisfying (2). Since every solution satisfies (2) for some θ, this means simply that

$$y'' + (\lambda_0 + f(x))y = g(x) \tag{7}$$

has a solution of class (L^2) whenever $g(x)$ is of class (L^2). This leads, however, to a contradiction.

In fact, Theorem (iv) in [4] implies that, if $f(x)$ is a continuous function satisfying the limsup-assumption of (I) and having the property that (4) holds for every solution of (5), then (7) cannot have a solution of class (L^2) for every choice of a continuous $g(x)$ of class (L^2) unless the following condition is fulfilled: Every solution $y = y(x)$ of (5) and every continuous $g = g(x)$ of class (L^2) determine a constant $a = a(y, g)$ for which the function

$$a - \int_0^x y(t)g(t)\,dt$$

becomes of class (L^2). But it will now be shown that such an $a = a(y, g)$ cannot always exist in the present case. This will complete the proof of (I).

Suppose that $a = a(y, g)$ always exists. Keep the (arbitrary) solution $y = y(x)$ of (5) fixed and, starting with an unspecified continuous $g(x)$ of class (L^2), replace $g(x)$ by the function $|g(x)| \operatorname{sgn} y(x)$. Since the latter function is again continuous and of class (L^2), the function defined by the last formula line must become of class (L^2) after the replacement, if the value of the constant a is altered in a suitable way. Accordingly, there belongs

to every continuous $g(x)$ of class (L^2) (and to the fixed solution $y(x)$ of (5)) a constant b for which the function

$$b - \int_0^x |y(t)g(t)|\,dt$$

becomes of class (L^2). But the latter function is monotone and cannot, therefore, be of class (L^2) unless it tends to 0 as $x \to \infty$. Consequently,

$$b = \int_0^\infty |y(t)g(t)|\,dt.$$

Accordingly,

$$\int_0^\infty y(x)g(x)\,dx$$

must exist, as a Lebesgue integral, for every continuous $g(x)$ of class (L^2). It must therefore exist for every $g(x)$ of class (L^2) (in fact, the continuous functions of class (L^2) are dense on the space of all functions of class (L^2)). In view of the well-known "converse of Schwarz's inequality," this is possible only if the fixed "kernel," $y(x)$, occurring in the last formula line, is of class (L^2). But $y(x)$ was chosen as an arbitrary solution of (5). Hence, every solution of (5) is of class (L^2). This means that (1) is of the *Grenzkreis* type and contradicts, therefore, the first assertion of (I).

The proof of (I) is now complete.

Proof of (II). Suppose that (3) satisfies the assumptions of (II). Since the solutions $y_1(x)$, $y_2(x)$ occurring in these assumptions have a non-vanishing Wronskian, the latter can be chosen to be any non-vanishing constant, and it is clear that, without violating the assumptions of (II), it can be supposed that

$$y'_1(0)y_2(0) - y'_2(0)y_1(0) = 1. \tag{8}$$

Then, since the Wronskian does not depend on x,

$$y'_1(x)y_2(x) - y'_2(x)y_1(x) = 1. \tag{9}$$

Since $y_1(x)$ and $y_2(x)$ are solutions of (5), two differentiations show that, (9) being assumed, the function

$$\int_0^x \{y_1(x)y_2(t) - y_2(x)y_1(t)\}g(t)\,dt \tag{10}$$

is a solution, $y(x)$, of (7) (for every continuous $g(x)$), and that both $y(0)$ and $y'(0)$ vanish for this $y(x)$. Consequently, the general solution of (7) results if (10) is added to the general solution, $c_1y_1(x) + c_2y_2(x)$, of (5), and the initial values $(x = 0)$ of the resulting solutions and their first derivatives will be the same for (7) as for (5); namely,

$$c_1y_1(0) + c_2y_2(0) \text{ and } c_1y'_1(0) + c_2y'_2(0)$$

for both (5) and (7).

Since (II) assumes that $y_2(x)$ and $g(x)$ are of class (L^2), and since the product of two functions of class (L^2) is of class (L^1),

$$c_1 = -\alpha \text{ and } c_2 = 0,$$

where

$$\alpha = \int_0^\infty y_2(x)g(x)\,dx, \tag{11}$$

is a possible choice of c_1 and c_2. Clearly, this choice reduces the preceding initial values to $-\alpha y_1(0)$ and $-\alpha y'_1(0)$. It is also seen from the last two formula lines that $c_1y_1(x) + c_2y_2(x)$ plus the function (10) can be contracted into $-g^*(x)$, if $g^*(x)$ is defined by (6). Accordingly,

$$y^*(x) = -g^*(x) \tag{12}$$

is a solution of (7) and belongs to the initial values

$$y^*(0) = -\alpha y_1(0), \qquad y^{*\prime}(0) = -\alpha y'_1(0), \tag{13}$$

if $g^*(x)$ denotes the function (6), and α the constant (11).

This fact will now be combined with the following remark: If c is an arbitrary constant, and if

$$y_1(x), \qquad y_2(x) \tag{14}$$

is a pair of solutions of (5) which satisfy the assumptions of (II), then

$$y_1(x) + cy_2(x), \qquad y_2(x) \tag{15}$$

is another pair of such solutions of (5). In fact, if (14) is replaced by (15) in (6), the resulting modification of the function (6) is seen to consist of the additional term

$$2cy_2(x)\int_x^\infty y_2(t)g(t)\,dt.$$

Hence, it is sufficient to ascertain that this additional term is a function of class (L^2) whenever $g(x)$ is. But (II) assumes that $y_2(x)$ is of class (L^2). Consequently, it is sufficient to ascertain that the integral multiplying $y_2(x)$ in the last formula line (exists and) is a bounded function of x. This, however, is obvious, since both $y_2(x)$ and $g(x)$ are of class (L^2), and so their product is of class (L^1).

Accordingly, (14) in (II) can be replaced by (15), where c is arbitrary. On the other hand, since the replacement of (14) by (15) is equivalent to the substitution

$$y_1(0) \to y_1(0) + cy_2(0),\ \ y'_1(0) \to y'_1(0) + cy_2(0);$$
$$y_2(0) \to y_2(0),\ y'_2(0) \to y'_2(0),$$

and since this substitution clearly is such as to leave the determinant (8) unaltered, (9) remains true (for every x, the Wronskian of two solutions of (5) being independent of x). Finally, it is seen that the substitution defined by the last formula line transforms (13) into

$$(16) \qquad y^*(0) = -\alpha y_1(0) - \alpha c y_2(0), \qquad y^{*\prime}(0) = -\alpha y'_1(0) - \alpha c y'_2(0).$$

It follows that, if $g(x)$ is any continuous function of class (L^2), and if (5) satisfies the assumptions of (II), then (7) has a solution $y = y^*(x)$ of class (L^2) satisfying (16). In fact, the assumptions of (II) require that (6) be a function of class (L^2), and so the assertion follows from (12) and from the fact that (14) in (6) can be replaced by (15).

It should be emphasized that the constant α occurring in (6) is independent of the arbitrary constant c. This follows from the circumstance that the replacement of (14) by (15) alters $y_1(x)$ only, whereas the definition, (11), of α contains $y_2(x)$ only.

It is seen from (16) that the solution $y^*(x) = y^*(x; c)$ of (7), a solution which is of class (L^2) for every c, satisfies the boundary condition (2) (belonging to a fixed θ) if and only if c satisfies (with reference to that θ) the condition

$$(17) \qquad \alpha\{y_1(0)\cos\theta + y'_1(\theta)\sin\theta\} + \alpha c\{y_2(0)\cos\theta + y'_2(0)\sin\theta\} = 0,$$

where the four constants $y_i(0)$, $y'_i(0)$ are those belonging to the pair, (14), of solutions of (5) which have been chosen originally (in other words, the four values $y_i(0)$, $y'_i(0)$ occurring in (17) are independent of both θ and c).

Clearly, there exist values of θ for which the linear equation (17) can be satisfied by a certain $c = c(\theta)$. Hence, there exist values of θ corresponding to which the inhomogeneous differential equation (7) has, for

every continuous $g(x)$ of class (L^2), a solution of class (L^2) satisfying the boundary condition (2). This means that there exist values of θ corresponding to which the spectrum belonging to (1) and (2) does not contain the particular parameter value, $\lambda = \lambda_0$, occurring in (5) and (7). Since the set of λ-values consisting of the continuous spectrum and of the cluster points of the point spectrum is independent of the choice of θ in (2), it follows that λ_0 cannot be cluster point of the spectrum. This completes the proof of (II).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

[1] O. Toeplitz, "Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen. I. Teil: Theorie der L-Formen," *Mathematische Annalen*, vol. 70 (1911), pp. 351-376.

[2] H. Weyl, "Singuläre Integralgleichungen," *ibid.*, vol. 66 (1908), pp. 273-324.

[3] ———, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *ibid.*, vol. 68 (1909), pp. 220-269.

[4] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

[5] ———, "Stability and spectrum in the wave mechanics of lattices," *Physical Review*, vol. 72 (1947), pp. 81-82.

MEMOIR ON ELLIPTIC DIVISIBILITY SEQUENCES.*

By Morgan Ward.

I. Introduction.

1. By an elliptic divisibility sequence we mean a sequence of integers,[1]

$$(h): \quad h_0, h_1, h_2, \cdots, h_n, \cdots$$

which is a particular solution of

$$(1.1) \qquad \omega_{m+n}\omega_{m-n} = \omega_{m+1}\omega_{m-1}\omega_n{}^2 - \omega_{n+1}\omega_{n-1}\omega_m{}^2$$

and such that h_n divides h_m whenever n divides m. Simple instances of such sequences are:

$$(1.2) \qquad h_n = n;$$

$$(1.3) \qquad h_n = (n/3)$$

where (n/p) is Legendre's symbol;

$$(1.4) \qquad h_n = (-8/n)$$

where (d/n) is Kronecker's symbol.[2]

$$(1.5) \qquad h_n = Q^{(1/2)-(n/2)}U_n$$

where $P = a + b$, $Q = ab$ and

$$(1.6) \qquad U_n = (a^n - b^n)/(a - b)$$

is a polynomial in P and Q satisfying a linear recurrence of order two. This polynomial is one of the two[3] fundamental numerical functions studied by Edouard Lucas in the first volume of this Journal (Lucas [1], [2]. See also Lucas [3]). Lucas continually emphasized the connections between his

* Received February 12, 1947.

[1] The case when the h are rational is not essentially more general.

[2] See Landau, *Vorlesungen*, I, p. 51. In the present case, $(-8/n) = 0$ if n is even and $(-8/n) = (-1)^{[n/4]}$ if n is odd, where $[n/4]$ denotes the greatest integer in $n/4$.

[3] The other function $V_n = a^n + b^n$ does not lead to a solution of (1.1) despite Lucas' assertion to this effect (Lucas [1], p. 203). See Bell [1]. We assume that P and Q are chosen so that (1.6) is an integer; for example, P an integer and Q plus or minus one.

numerical functions and the trigonometric functions, and claimed to have made a remarkable generalization connecting numerical functions defined by a linear recurrence of order three or four with the elliptic functions. (See Bell [1] for a review and evaluation of Lucas' claims. Lucas apparently published nothing on the subject save scattered hints.)

Since (1. 1) is the fundamental relation on which the real multiplication theory of elliptic functions rests,[4] a systematic study of (h)-sequences should throw some light on Lucas' conjecture. In addition (h)-sequences are of arithmetical interest on their own account; they appear to be the simplest type of non-linear [5] divisibility sequence, and yet most of the properties of Lucas' linear (U)-sequence carry over to them. The investigations which follow show conclusively that if any such generalization as Lucas conjectured exists, it must be looked for in the direction of the complex multiplication theory of elliptic functions. The arithmetical properties of elliptic divisibility sequences turn out to be quite different from those of numerical functions defined by linear recurrences or order greater than two.[6]

2. The main results of the memoir are as follows. We may confine ourselves to sequences in which $h_0 = 0$, $h_1 = 1$ and not both h_2 and h_3 vanish.[7]

A solution of (1. 1) satisfying these conditions is an elliptic divisibility sequence if and only if h_2, h_3, h_4 are integers and h_2 divides h_4. Every such solution is uniquely determined by the initial values of h_2, h_3 and h_4 and may be parameterized by elliptic functions provided that h_2 and h_3 are not zero.[8] The invariants g_2 and g_3 of the associated $\wp$ function are rational functions of h_2, h_3 and h_4.

Every divisibility sequence with $h_2 = 0$ is essentially equivalent to the solution (1. 4) of the previous section and every rational solution of (1. 1) with $h_0 = 0$, $h_1 = 1$ and not both h_2 and h_3 zero is essentially equivalent to an integral elliptic divisibility sequence.

(h) reduces essentially to the solution (1. 2) of Section 1 if and only if

[4] $\omega_n = \sigma(nu)/\sigma(u)^{n^2}$ satisfies (1. 1) where $\sigma(z)$ is the Weierstrass sigma function.

[5] A divisibility sequence is said to be linear if it satisfies a linear recurrence relation. See Hall [1].

[6] The theory of such functions was initiated by Carmichael. (Carmichael [1]). See also Ward [1] and the references given there.

[7] If both h_2 and h_3 vanish, there exist integral solutions of (1. 1) which are not divisibility sequences and which are not determined by any fixed number of initial values. These and other special sequences are discussed in Chapter VII.

[8] If h_2 or h_3 is zero, h_n is trivially a product of powers of $\pm h_3$ or $\pm h_2$ and h_4. (This case is discussed in Chapter VII).

g_2 and g_3 both vanish, and (h) reduces essentially to Lucas' solutions (1. 5) if and only if neither g_2 nor g_3 vanishes, but the elliptic discriminant $g_2{}^3 - 27g_3{}^2$ vanishes.

An integer m is said to be a divisor of (h) if it divides some term h_k with $k > 0$. If m divides h_k but does not divide h_l when l divides k, then k is called a rank of apparition of m in (h).[9] Every prime p which does not divide both h_3 and h_4 has precisely one rank of apparition ρ, and (h) is periodic modulo p with period $\rho\tau$ where τ is a certain arithmetical function of p and (h) which can be exactly determined.[10] Similar results hold for a composite modulus m.

If the least positive residues modulo m of the successive values $U_0, U_1, U_2, \cdots$ of any Lucas function are calculated, the pattern of residues exhibits interesting symmetries.[11] These symmetries extend to elliptic sequences, and find their ultimate explanation in the periodicity of the second kind of the Weierstrass sigma function.

3. The plan of the paper is sufficiently indicated by the chapter titles. We develop first those arithmetrical properties of the sequences which can be proved without the use of elliptic functions; the important modular periodicity, however, depends on the elliptic function representation. Our conclusions regarding Lucas' conjectures are given in the final chapter.

The terminology describing the arithmetical properties of the elliptic sequences is chosen to agree with that used for linear sequences (Hall [1], Ward [2]). We use the standardized arithmetical notations of Landau's *Vorlesungen*; in particular, if $a, b, \cdots$ are integers or ideals, $a \mid b$ for "a divides b" and $(a, b, \cdots)$ for the greatest common divisor of $a, b, \cdots$. We denote the least common multiple of $a, b, \cdots$ by $[a, b, \cdots]$.

The results of elliptic function theory which are used in Chapter IV may be found in any standard text; the account of (1. 1) in Halphen's Treatise is particularly complete, and many of his results may be restated as theorems about elliptic sequences.

---

[9] This definition is due to M. Hall. See Hall [1].

[10] In the terminology of the theory of linear recurrences, ρ is the "restricted period" of (h) modulo p. See Carmichael [1].

[11] Typical examples are given by the residues of the Fibonacci sequence $0, 1, 1, 2, 3, \cdots$ for small integral moduli. See also the table of elliptic sequences modulo three in Section 7 of Chapter III.

II. Elementary Properties of Sequences.

4. We shall confine ourselves in the next five chapters to sequences whose first two initial values are zero and one. If in addition neither the third nor the fourth value vanishes, we call the sequence "general." Sequences which violate one or more of these restrictions are called "special," and are discussed in detail in Chapter VII. It turns out that the only sequences (h) which have any arithmetical interest satisfy the following conditions:

$$h_0 = 0,\ h_1 = 1;\ \text{not both } h_2 \text{ and } h_3 \text{ zero.} \tag{4.1}$$

We call such sequences "proper." Proper sequences include as well as the general sequence, two special sequences in which either $h_2 = 0$, $h_3 \neq 0$ or $h_2 \neq 0$, $h_3 = 0$. We shall begin by proving the following basic theorem.

THEOREM 4.1. *Let (h) be a proper solution of (1.1) so that (4.1) holds and also*

$$h_{m+n}h_{m-n} = h_{m+1}h_{m-1}h_n^2 - h_{n+1}h_{n-1}h_m^2, \qquad m \geqq n \geqq 1. \tag{4.11}$$

Then (h) is an elliptic divisibility sequence if and only if

$$h_2, h_3 \textit{ and } h_4 \textit{ are integers;} \tag{4.2}$$

$$h_2 \textit{ divides } h_4. \tag{4.3}$$

Furthermore, the sequence (h) is uniquely determined by the three initial values h_2, h_3 and h_4.

Proof. Assume first that

$$h_2 \neq 0. \tag{4.31}$$

Since the necessity of the conditions (4.2) and (4.3) is evident, assume conversely that (h) is a solution of (1.1) for which (4.1), (4.2), (4.3) and (4.31) all hold. We shall first prove by induction that all terms of (h) are integers and that h_2 divides h_{2n}. We then make a second induction to prove that (h) is a divisibility sequence; that is

$$h_r \text{ divides } h_s \text{ if } r \text{ divides } s. \tag{4.4}$$

A third and final induction shows that if n is greater than four, h_n is uniquely determined if $h_0, h_1, \cdots, h_{n-1}$ are uniquely determined.

We obtain the following important formulas from (4.11) on taking first $m = n + 1$, $n = n$ and then $m = n + 1$ and $n = n + 1$:[12]

$$h_{2n+1} = h_{n+2}h^3{}_n - h_{n-1}h^3{}_{n+1}, \qquad n \geqq 1. \tag{4.5}$$

$$h_{2n}h_2 = h_n(h_{n+2}h^2{}_{n-1} - h_{n-2}h^2{}_{n+1}), \qquad n \geqq 2. \tag{4.6}$$

We begin the first induction by assuming that

(α) $h_0, h_1, \cdots, h_{n-1}$ *are integers;*

(β) h_2 *divides* h_{2r}, $2r < n$; $n \geqq 5$.

If n is odd, say $n = 2k + 1$, we conclude from (α) and (4.5) with $n = k$ that h_n is an integer. If n is even, say $n = 2k$, then $k \geqq 3$ and (4.6) gives

$$h_nh_2 = h_k(h_{k+2}h^2{}_{k-1} - h_{k-2}h^2{}_{k+1}). \tag{4.7}$$

Since $k + 2 < 2k$ and the suffices $k \pm 2$, $k \mp 1$ are of opposite parity, $h_{k+2}h^2{}_{k-1} - h_{k-2}h^2{}_{k+1}$ is an integer divisible by h_2. Hence h_n is an integer. But if k is even, h_k is divisible by h_2 and if k is odd, $h^2{}_{k-1}$ and $h^2{}_{k+1}$ are divisible by $h^2{}_2$. Hence in either case h_n is divisible by h_2. The first part of the theorem follows then by induction on n.

We prove (4.4) by a second induction. Assume that

(γ) $h_r \mid h_s$ provided that $r \mid s$ for $r \leqq s < n$.

We observe that (γ) is true if $n \leqq 5$. Hence we may assume that $n > 5$. Now consider h_n, and suppose that $n = uv$. We wish to show that $h_u \mid h_{uv}$ and it is evidently allowable to assume that $u \geqq 2$ and $v \geqq 2$.

Suppose first that $h_u \neq 0$. Then if v is even, (4.6) gives

$$h_{uv}h_2 = h_{uv/2}\,(h_{(uv/2)+2}h^2{}_{(uv/2)-1} - h_{(uv/2)-2}h^2{}_{(uv/2)+1}).$$

The parenthesis is divisible by h_2. But by (γ), $h_u \mid h_{uv/2}$. Hence $h_u \mid h_v$.

If v is odd, u and uv are of the same parity. Hence on taking $m + n = uv$ and $m - n = u$ in (1.1) we obtain the relation

$$h_{uv}h_u = h_{m+1}h_{m-1}h^2{}_n - h_{n+1}h_{n-1}h^2{}_m$$

Since $m = u(v + 1)/2$ and $n = u(v - 1)/2$, we conclude from (γ) that the right side of this expression is divisible by $h^2{}_u$. Hence since $h_u \neq 0$, $h_u \mid h_{uv}$.

---

[12] As will be evident, if we define a sequence (h) recursively by (4.5) and (4.6) (taking $h_0 = 0$, $h_1 = 1$, $h_2 \neq 0$ and h_3, h_4 arbitrary) then conversely we obtain a solution of (1.1).

Now assume that $h_u = 0$. We shall need the following lemma which will be important in other connections. We shall postpone its proof until we have completed the proof of the theorem.

LEMMA 4.1. *Let* (h) *be any solution whatever of* (1.1) *with initial values* $h_0 = 0$, $h_1 = 1$ *and not both* h_2 *and* h_3 *zero. Then if two consecutive terms of* (h) *vanish, all terms of* (h) *vanish beyond the third.*[13]

Since $h_u = 0$, $h_{u(v-1)}$ is zero by (γ). Then on taking $m = uv$ and $n = u$ in (4.11), we obtain

$$h_{u(v+1)}h_{u(v-1)} = h_{uv+1}h_{uv-1}h^2{}_u - h_{u+1}h_{u-1}h^2{}_{uv}$$

Hence either $h_{uv} = 0$ or $h_{u+1}h_{u-1} = 0$ so that two consecutive terms of (h) vanish. Then by the lemma, $h_3 = h_4 = h_5 \cdots = h_{uv} = 0$. Hence in all cases, h_u divides h_{uv}. (4.4) now follows by induction on n in (γ).

Finally, the unicity of (h) follows directly from the formulas (4.5) and (4.6) by a brief induction.

It remains to discuss the case when h_2 vanishes. Then $h_3 \neq 0$ and we shall prove in Theorem 23.1 of Chapter VII that the general term of (h) is as follows:[14]

$$h_n = \begin{matrix} 0, \ n \text{ even}; \\ (-1)^{[n/4]}h_3{}^{(n^2-1)/8}, \ n \text{ odd}, \ h_3 \neq 0. \end{matrix} \tag{23.4}$$

Hence Theorem 4.1 holds in this special case, too, completing the proof.

The lemma is proved as follows. If two consecutive terms of (h) vanish, then two consecutive terms of smallest suffix vanish; let them be h_r and h_{r+1}. Then $r \geqq 3$, and in the interval $0 < n < r$, not both h_n and h_{n+1} are zero. I say that $h_n \neq 0$ in this interval. For if $r = 3$, h_1 and h_2 are not zero. Assume then that $r > 3$, and $h_k = 0$. Then $2 \leqq k \leqq r-2$ and by the minimal property of r,

$$h_{k-1}h_{k+1} \neq 0. \tag{4.8}$$

Now if $k < r/2$, choose l so that $l + k = r$ and take $m = l$ and $n = k$ in (4.11). Since $h_k = h_r = 0$, we obtain $-h_{k-1}h_{k+1}h^2{}_l = 0$. Hence by (4.8), $h_l = 0$. Now replace l by $l+1$. Since $h_{r+1} = h_k = 0$, we obtain

[13] It is shown in Chapter VII that this lemma is not necessarily true if both h_2 and h_3 are zero. If, however, (h) is assumed to be a divisibility sequence, the vanishing of h_2 and h_3 entails the vanishing of all subsequent terms.

[14] If $h_2 = 0$, $h_3 = 1$ we obtain the particular periodic solution $h_n = (-8/n)$ mentioned in the Introduction.

$-h_{k-1}h_{k+1}h^2_{l+1} = 0$. Hence $h_{l+1} = 0$. But $h_l = h_{l+1} = 0$ contradicts the minimal property of r. Next, if $k = r/2$, we take $l = k + 1$ and find that $h_{k+1} = 0$, contrary to (4.8). If $k > r/2$, we take $k + l = r$ and take $m = k$ and $n = l$ in (4.11), obtaining as before $h_l = 0$. Then replacing l by $l + 1$, we find $h_{l+1} = h_l = 0$ contradicting again the minimal property of r. Hence we may assume

$$h_r = 0, \quad h_{r+1} = 0, \quad h_n \neq 0, \qquad 0 < n < r. \tag{4.9}$$

I say that $r = 3$. For if $r > 3$, $h_3 \neq 0$ by (4.9). Hence on taking $m + n = 2r - 3$ and $m - n = 3$ in (4.11), we obtain the relation

$$h_{2r-3}h_3 = h_{r+1}h_{r-1}h^2{}_{r-3} - h_{r-4}h_{r-2}h^2{}_r$$

where all the suffices are $\geqq 0$. Hence by (4.9), $h_{2r-3} = 0$. But by (4.5) with $n = r - 2$,

$$0 = h_{2r-3} = h_rh^3{}_{r-2} - h_{r-3}h^3{}_{r-1} = -h_{r-3}h^3{}_{r-1}.$$

Hence either h_{r-1} or h_{r-3} is zero, contradicting (4.9).

But if $r = 3$, then $h_2 \neq 0$ but $h_3 = h_4 = 0$ and we find by a brief induction from (4.5) and (4.6) that $h_n = 0$ for $n \geqq 3$. This completes the proof of the lemma.

5. An integer m is said to be a divisor of the sequence (h) if it divides some term with positive suffix. If m divides h_ρ but does not divide h_r if r divides ρ, then ρ is called a rank of apparition of m in (h).

THEOREM 5.1. *An elliptic divisibility sequence admits every prime p as a divisor. Furthermore, p has at least one rank of apparition smaller than $2p + 2$.*

Proof. If none of $h_1, h_2, \cdots, h_{p+1}, h_{p+2}$ is divisible by p, each of the p numbers

$$\frac{h_{r-1}h_{r+1}}{h_r{}^2}, \qquad (r = 2, 3, \cdots, p+1)$$

is congruent modulo p to one of the numbers $1, 2, \cdots, p-1$. Hence at least two are congruent to one another; say

$$\frac{h_{n-1}h_{n+1}}{h_n{}^2} \equiv \frac{h_{m+1}h_{m-1}}{h_m{}^2} \equiv c \pmod p$$

when $2 \leqq n < m \leqq p + 1$ and c is an integer. But then (4.11) gives the congruence

$$h_{m+n}h_{m-n} \equiv 0 \pmod p.$$

Since $m-n<p+2$ and p is prime, h_{m+n} is divisible by p. Hence the smallest rank of apparition of p is at most $m+n$ and hence less than or equal to $2p+1$. $2p+1$ is the best upper bound possible. For if $p=2$ and $h_0=0, h_1=h_2=h_3=h_4=1$, then $\rho=5$.

THEOREM 5.2. *Let p be a prime divisor of an elliptic sequence (h), and let ρ be its smallest rank of apparition. Then if*

$$(5.1)\qquad h_{\rho+1}\not\equiv 0 \pmod p,$$

$$(5.2)\qquad h_n\equiv 0 \pmod p \text{ if and only if } n\equiv 0 \pmod \rho.$$

Proof. By the definition of ρ,

$$(5.3)\qquad h_\rho\equiv 0 \pmod p,\qquad h_r\not\equiv 0 \pmod p,\qquad 0<r<\rho.$$

Since (h) is a divisibility sequence h_ρ divides h_n if ρ divides n. Hence $h_n\equiv 0 \pmod p$ if $n\equiv 0 \pmod \rho$. We prove the converse by mathematical induction. Assume that (5.2) holds for $n<k$. We can clearly assume that $k\geqq \rho+2$. Consider h_k. If $h_k\not\equiv 0 \pmod p$, we cannot have $k\equiv 0 \pmod \rho$. Hence (5.2) will then hold for $n<k+1$. If $h_k\equiv 0 \pmod p$, divide k by ρ and let the quotient be q and the remainder r:

$$(5.4)\qquad k=q\rho+r,\qquad 0\leqq r<\rho.$$

We shall show that the assumption that $r>0$ in (5.4) leads to a contradiction. (5.2) then immediately follows by mathematical induction on k.

Assume then that $r>0$ in (5.4). Then taking $m=q\rho$ and $n=r$ in (4.11) $h_{m+n}\equiv h_m\equiv 0$. Hence we obtain the congruence

$$h_{q\rho+1}h_{q\rho-1}h^2{}_r\equiv 0 \pmod p.$$

Now $q\rho-1<k$ and $q\rho-1$ is not divisible by ρ. Hence $h_{q\rho-1}\not\equiv 0 \pmod p$ by the hypothesis of the induction.

Also $h_r\not\equiv 0 \pmod p$ by (5.3). Hence since p is a prime,

$$h_{q\rho+1}\equiv 0 \pmod p.$$

If $q=1$, (5.1) is contradicted. If $q=2$, then on taking $m=\rho$ in (4.5), we find that $h_{q\rho+1}=h_{\rho+2}h^3{}_\rho-h_{\rho+1}h^3{}_{\rho-1}$. Hence

$$h_{\rho+1}h^3{}_{\rho-1}\equiv 0 \pmod p.$$

Since p is a prime, either $h_{\rho+1}\equiv 0 \pmod p$ or $h_{\rho-1}\equiv 0 \pmod p$, contradicting (5.1) or (5.3). Hence $q>2$. Now take $m=(q-1)\rho$ and

$n = \rho + 1$ in (1.2). Then $m - n > 0$ and since $h_{m+n} \equiv h_m \equiv 0 \pmod{p}$, we obtain the congruence

$$h_{(q-1)\rho+1} h_{(q-1)\rho-1} h^2_{\rho+1} \equiv 0 \pmod{p}.$$

But since $0 < (q-1)\rho - 1 < (q-1)\rho + 1 < q\rho + 1 \leqq k$, $h_{(q-1)\rho-1}$, and $h_{(q-1)\rho+1}$ are incongruent to zero modulo p. Hence since p is a prime, $h_{\rho+1} \equiv 0 \pmod{p}$, contradicting (5.1). Hence r must be zero in (5.4), and the proof is complete.

6. The following theorem is a companion to Theorem 5.1.

THEOREM 6.1. *Let p be a prime divisor of an elliptic sequence (h), and let ρ be its smallest rank of apparition. If*

$$h_{\rho+1} \equiv 0 \pmod{p} \tag{6.1}$$

then $\rho \leqq 3$ and

$$h_n \equiv 0 \pmod{p}, \qquad n \geqq \rho. \tag{6.2}$$

Proof. By definition of ρ,

$$h_\rho \equiv 0 \pmod{p}, \qquad h_r \not\equiv 0 \pmod{p}, \qquad 0 < r < \rho. \tag{6.3}$$

We shall show first that the assumption

$$\rho > 3 \tag{6.4}$$

leads to a contradiction with (6.1) and (6.3). If (6.4) holds, $h_3 \not\equiv 0 \pmod{p}$. Taking $m + n = 2\rho - 2$ and $m - n = 3$ in (4.11), we obtain the relation

$$h_{2\rho-3} h_3 = h_{\rho+1} h_{\rho-1} h^2_{\rho-3} - h_{\rho-4} h_{\rho-2} h^2_\rho$$

where all the suffixes are $\geqq 0$ by (6.4). Thus since p is a prime,

$$h_{2\rho-3} \equiv 0 \pmod{p}.$$

Now taking $n = \rho - 2$ in the relation (4.5), we find that $h_{2\rho-3} = h_\rho h^2_{\rho-2} - h_{\rho-3} h^2_{\rho-1}$. Hence $h_{\rho-3} h^3_{\rho-1} \equiv 0 \pmod{p}$. Since p is a prime, either $h_{\rho-1}$ or $h_{\rho-3}$ is divisible by p contrary to (6.3). Hence $\rho \leqq 3$. If $\rho = 2$, $h_{2n} \equiv 0 \pmod{p}$ since h_2 divides h_{2n}. Since $h_3 = 0$ by (6.1), $h_5 \equiv 0$ by (4.5) with $n = 2$. It is now easy to prove by induction from (4.5) that $h_{2n+1} \equiv 0 \pmod{p}$. Hence if $\rho = 2$, $h_n \equiv 0 \pmod{p}$, for $n \geqq \rho$.

If $\rho = 3$, $h_2 \not\equiv 0 \pmod{p}$ and we easily prove from (4.5), (4.6) and $h_3 \equiv h_4 \equiv 0 \pmod{p}$ that $h_n \equiv 0 \pmod{p}$ for $n \geqq 3$. This completes the proof of the theorem.

The following two theorems follow directly from Theorems 5.1 and 6.1.

THEOREM 6.2. *A necessary and sufficient condition that a prime p have exactly one rank of apparition in an elliptic sequence (h) is that p is not a common divisor of h_3 and h_4.*[15]

THEOREM 6.3. *A necessary and sufficient condition that every prime have precisely one rank of apparition in an elliptic sequence (h) is that h_3 and h_4 have no common factor.*

The following theorem now follows from a known result: (Ward [3])

THEOREM 6.4. *If (h) is an elliptic sequence in which the initial values h_3 and h_4 are co-prime, then $(h_n, h_m) = h_{(m,n)}$.*

III. The Numerical Periodicity and Symmetry Modulo p of Sequences.

7. A sequence (s) of rational integers is said to be numerically periodic modulo m if there exists a positive integer π such that

$$s_{n+\pi} \equiv s_n \pmod{m} \tag{7.1}$$

for all sufficiently large n. If (7.1) holds for all n, then (s) is said to be purely periodic modulo m. The smallest such integer π for which (7.1) is true is called the period of (h) modulo m. All other periods are multiples of it.

We shall show in this chapter that any elliptic sequence is numerically periodic for any prime modulus and purely periodic for all primes which do not divide both h_3 and h_4. The culminating result is the following theorem which shows precisely how the period and rank are connected.

THEOREM 11.1. *Let (h) be an elliptic divisibility sequence and p an odd prime whose rank of apparition ρ is greater than three. Let e be an integral solution of the congruence*

$$e \equiv h_2/h_{\rho-2} \pmod{p}, \tag{11.1}$$

and let ϵ and κ be the exponents to which e and $h_{\rho-1}$ respectively belong modulo p. *Then (h) is purely periodic* modulo p, *and its period π is given by the formula $\pi = \tau\rho$ where*

$$\tau = 2^{\alpha}[\epsilon, \kappa]. \tag{7.11}$$

---

[15] Since h_2 divides h_4, a common divisor of h_2 and h_3 is a common divisor of h_3 and h_4.

Here $[\epsilon, \kappa]$ *is the least common multiple of* ϵ *and* κ *and the exponent* α *is determined as follows:*

$\alpha = +1$ *if and only if* ϵ, κ *are both odd,*

$\alpha = -1$ *if and only if* ϵ, κ *are both even and both divisible by precisely the same power of* 2;

$\alpha = 0$ *in all other cases.*

I have been unable to establish the numerical periodicity of (h) sequences by elementary means; that is, without the use of their elliptic function representation. It turns out that the two invariants g_2 and g_3 of the elliptic function associated with this representation are each expressible as a polynomial in h_2, h_3 and h_4 with integral coefficients divided by a product of powers of h_2, h_3, two and three.[16] The arithmetical consequences of the elliptic function representation do not therefore apply to the primes two and three, or more generally to any prime dividing h_2 or h_3. We shall begin by discussing these exceptional primes.

There are eight *a priori* possible types of elliptic sequences modulo two distinguished by the possible residues of h_2, h_3 and h_4 modulo two. But since h_2 divides h_4, sequences with $h_2 \equiv 0 \pmod 2$ and $h_4 \equiv 1 \pmod 2$ cannot occur. The six possibilities which are left are listed in the following table..

ELLIPTIC SEQUENCES MODULO TWO

Type Number	Residues of h_i modulo two						Rank ρ	Period π
	h_0	h_1	h_2	h_3	h_4	h_5		
1	0	1	0	0	0	0	2	1
2	0	1	1	0	0	0	3	1
3	0	1	0	1	0	1	2	2
4	0	1	1	0	1	1	3	3
5	0	1	1	1	0	1	4	4
6	0	1	1	1	1	0	5	5

Theorem 11.1 however is not true for the two types five and six for which ρ is greater than three. In both cases $\epsilon = \kappa = 1$ so that the formula (7.11) gives $\pi = 2\rho$ instead of $\pi = \rho$. Thus the restriction to odd primes is necessary.

The twenty-one possible types of sequences modulo three are listed below. In each case when the rank ρ is greater than three, ϵ and κ are listed and also the multiplier $\tau = 2^{\alpha}[\epsilon, \kappa]$. The ranks and periods were obtained by direct computation for each type from the formulas (4.5) and (4.6) taken modulo three. The table thus shows that Theorem 11.1 is true if $p = 3$.

---

[16] See Chapter IV, Section 13, formulas (13.6) and (13.7).

ELLIPTIC SEQUENCES MODULO THREE

Type No.	h_0	h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9	h_{10}	h_{11}	h_{12}	h_{13}	h_{14}	Rank	Period	Exponents ϵ	κ	τ
1	0	1	0	0	0											2	1			
2	0	1	1	0	0											3	1			
3	0	1	2	0	0											3	1			
4	0	1	0	1	0	2	0	2								2	8			
5	0	1	0	2	0	1	0	2								2	4			
6	0	1	1	0	1	1	0	2	2	0	2	2				3	12			
7	0	1	1	0	2	2										3	6			
8	0	1	2	0	1	2										3	3			
9	0	1	2	0	2	1	0	2	1	0	1	2				3	12			
10	0	1	1	1	0	2	2	2								4	8	1	1	2
11	0	1	1	1	1	0	2	2	2	2						5	10	1	1	2
12	0	1	1	1	2	1	0	2	1	2	2	2				6	12	2	1	2
13	0	1	1	2	0	1	2	2								4	8	1	2	2
14	0	1	1	2	1	2	2	0	1	1	2	1	2	2		7	7	2	2	1
15	0	1	2	2	2	2	1	0	2	1	1	1	1	2		7	14	1	1	2
16	0	1	1	1	2	2	0									5	5	2	2	1
17	0	1	2	1	0	2	1	2								4	8	1	1	2
18	0	1	2	1	1	0	2	2	2	1	1					6	12	2	1	2
19	0	1	2	1	2	0										5	5	2	2	1
20	0	1	2	2	0	1	1	2								4	8	1	2	2
21	0	1	2	2	1	0	2	1	1	2	2					5	10	1	1	2

This table affords simple illustrations of the modular symmetry of sequences which was alluded to in the introduction. For example, consider type 14 and 15. For type 14, we have $h_{\rho-n} \equiv -h_n \pmod 3$. For type 15, $h_{\rho-n} \equiv h_n \pmod 3$; $h_{\rho+n} \equiv h_{\rho-n} \pmod 3$. We shall see that for primes other than two and three, the origin of this symmetry is the periodicity of the second kind of the elliptic sigma functions.

Now consider primes which divide the initial values h_2 and h_3. We have shown in Section **6** that primes which divide both h_3 and h_4 divide every subsequent term of (h). We call such primes "null divisors" of (h).[17] If p is a null divisor, then (h) is numerically periodic modulo p with the period one.[18] Since h_2 divides h_4, primes which divide both h_2 and h_3 are

---

[17] The terminology is borrowed from the theory of linear divisibility sequences. See Ward [2].

[18] The first types listed in the tables of elliptic sequences modulo two and modulo three afford simple illustrations. If (h) is a null sequence modulo p, it appears to be very difficult to specify the exact power of p dividing h_n given only the initial values of (h).

also null divisors. On excluding null divisors, we have as well as the "general case"

(7.2) $$h_2h_3 \not\equiv 0 \pmod{p},$$

two special cases:

(7.3) $$h_2 \equiv 0 \pmod{p}, \qquad h_3 \not\equiv 0 \pmod{p};$$

(7.4) $$h_3 \equiv 0 \pmod{p}, \qquad h_2 \not\equiv 0 \pmod{p}.$$

These cases are disposed of by the following theorem which is a simple consequence of the theorems on special sequences given in Chapter VII.

THEOREM 7.1. *If condition* (7.3) *holds, then*

$$h_{2n} \equiv 0 \pmod{p}, \qquad h_{2n+1} \equiv (-1)^{[n/4]} h_3^{(n^2-1)/8} \pmod{p}.$$

If condition (7.4) *holds, then*

$$h_{3n} \equiv 0 \pmod{p},$$
$$h_{3n+1} \equiv (-1)^{n(n-1)/2} h_2^{n(n-1)/2} h_4^{n(n+1)/2} \pmod{p},$$
$$h_{3n+2} \equiv -(-1)^{n(n+1)/2} h_2^{n(n+1)/2} h_4^{n(n-1)/2} \pmod{p}.$$

We see that in either case (h) is purely periodic modulo p. Its period depends in a simple way on the exponents to which its initial values belong modulo p.

8. The general case depends upon the following theorem which is proved in Chapter V, by the use of elliptic functions. All further developments in this chapter are obtained from this theorem by elementary means.

THEOREM 8.1. *Let p be a prime greater than three*[19] *which divides neither h_2 nor h_3. Then if ρ is its rank of apparition their exist two integers a and b such that*

(8.1) $$h_{\rho-n} \equiv a^n b h_n \pmod{p}, \qquad (n = 0, 1, 2, \cdots, \rho).$$

If we calculate successively the least positive residues modulo p of the first ρ terms of (h), the theorem states that there is a certain symmetry in the distribution of these residues. The theorems which follow not only lead to the proof of the periodicity of (h) modulo p, but also state symmetries in the pattern of least positive residues of successive blocks of ρ terms of (h). The final result of these symmetries is to determine the residues modulo p of

[19] The table of sequences modulo three shows that this theorem is also true if $p = 3$.

all terms of (h) in terms of the integers a and b of the theorem and the residues of the first $[\rho/2]$ terms. The next theorem shows how a and b may be determined modulo p.

THEOREM 8.2. *If a and b are the integers specified in Theorem* 8.1 *and if c is determined by the congruence*

$$ac \equiv 1 \pmod p \tag{8.2}$$

then the following congruences hold modulo p:

$$a \equiv h_{\rho-2}/h_2 h_{\rho-1};\quad b \equiv h_2 h^2_{\rho-1}/h_{\rho-2};\quad b \equiv h_{\rho-1}c. \tag{8.3}$$

$$a^\rho b^2 \equiv 1;\quad c^\rho \equiv b^2. \tag{8.4}$$

$$a^2 \equiv -h_{\rho-1}/h_{\rho+1};\quad b^2 \equiv -h_{\rho+1}h_{\rho-1}. \tag{8.5}$$

Proof. Let n successively equal 1 and $\rho - 1$, in (8.1). We obtain:

$$h_{\rho-1} \equiv ab \pmod p \tag{8.6}$$

and [20] $1 \equiv h_1 \equiv a^{\rho+1} b h_{\rho-1} \equiv a^\rho b^2$. (8.4) now follows and (8.6) and (8.2) imply that $b \equiv h_{\rho-1}c$ which is the last part of (8.3).

Next, put n equal to two in (8.1). Then

$$h_{\rho-2} \equiv a^2 b h_2 \equiv a h_{\rho-1} h_2 \pmod p,$$

the last step following from (8.6). This result is equivalent to the first part of (8.3). The second part follows now by (8.6). It remains to prove (8.5). Consider $h_{\rho+1}$. Assume first that ρ is odd:

$$\rho = 2\sigma + 1 \geqq 5. \tag{8.7}$$

Then on putting n equal to $\sigma + 1$ and σ in (4.6), we obtain

$$h_{\rho+1} = h_{\sigma+1} h_{\sigma+3} h^2_\sigma - h_{\sigma+1} h_{\sigma-1} h^2_{\sigma+2}, \tag{8.8}$$

$$h_{\rho-1} = h_\sigma h_{\sigma+2} h^2_{\sigma-1} - h_\sigma h_{\sigma-2} h^2_{\sigma+1}. \tag{8.9}$$

But by (8.1) and (8.7), the following congruences hold modulo p:

$$h_{\sigma+1} \equiv a^\sigma b h_\sigma;\quad h_{\sigma+3} \equiv a^{\sigma-2} b h_{\sigma-2};\quad h_\sigma \equiv a^{\sigma+1} b h_{\sigma+1};$$

$$h_{\sigma-1} \equiv a^{\sigma+1} b h_{\sigma+1};\quad h_{\sigma+2} \equiv a^{\sigma-1} b h_{\sigma-1}.$$

---

[20] The modulus p will be omitted here and elsewhere when no confusion can arise.

On substituting these expressions into (8. 8) and simplifying, (8. 9) gives the congruence

(8. 10) $$h_{\rho+1} \equiv -a^{2\rho-2}b^4 h_{\rho-1} \pmod p.$$

When ρ is even, this congruence may be shown to hold in essentially the same way.

Now by (8. 2) and (8. 4) Theorem 8. 2, $a^{2\rho}b^4 \equiv 1 \pmod p$. Hence (8. 10) implies that $h_{\rho+1} \equiv -a^{-2}h_{\rho-1} \pmod p$, and this congruence is equivalent to the first part of (8. 5). The second part of (8. 5) now follows by (8. 6), completing the proof.

9. The theorems of this section give the fundamental symmetries of (h) modulo p.

LEMMA 9. 1. *With the notation of Theorems* (8. 1) *and* (8. 2), *the following congruence is valid for all positive integers* n:

(9. 1) $$h_{\rho+n} \equiv -bc^n h_n \pmod p.$$

Proof. Assume first that $0 \leqq n \leqq \rho$. Since

$$h_{\rho+n}h_{\rho-n} = h_{\rho+1}h_{\rho-1}h_n^2 - h_{n+1}h_{n-1}h_\rho^2$$

and p divides h_ρ, we obtain from (8. 5) the congruence $h_{\rho+n}h_{\rho-n} \equiv -b^2h_n^2 \pmod p$ or by (8. 1), $h_{\rho+n}a^nbh_n \equiv -b^2h_n^2 \pmod p$. If $0 < n < \rho$, we may cancel bh_n. We then obtain (9. 1) on multiplying by c^n. Since the cases $n = 0$ and $n = \rho$ are trivially satisfied, (9. 1) *is true for* $0 \leqq n \leqq k\rho$ *if* k *equals one.*

We now proceed by induction on k. Suppose that (9. 1) is true for $0 \leqq n \leqq k\rho$ and assume that $k\rho \leqq n \leqq (k+1)\rho$. Then since

$$h_{n+\rho}h_{n-\rho} = h_{n+1}h_{n-1}h_\rho^2 - h_{\rho+1}h_{\rho-1}h_n^2$$

and p divides h_ρ, we obtain from (8. 5) the congruence

(9. 2) $$h_{n+\rho}h_{n-\rho} \equiv b^2h_n^2 \pmod p.$$

Now $0 \leqq n - \rho \leqq k\rho$. Hence by the hypothesis of the induction,

(9. 3) $$h_{n-\rho} \equiv -(a^{n-\rho}/b)h_n \pmod p.$$

Hence if $k\rho < n < (k+1)\rho$, (9. 2) and (9. 3) give the congruence $b^3h_n \equiv a^{n-\rho}h_{\rho+n} \pmod p$. Since $a^\rho b^2 \equiv 1$ by (8. 4) and $a^nc^n \equiv 1$ by (8. 1), this last congruence gives (9. 1) on multiplication by $a^\rho c^n$. Since (9. 1) holds

trivially for $n = k\rho$ or $n = (k+1)\rho$, and has been proven true for $0 \leqq n \leqq \rho$, the induction is completed.

THEOREM 9.2. *Under the hypothesis of Lemma* (9.1),

$$h_{k\rho+n} \equiv (-1)^k c^{kn} b^{k^2} h_n \pmod p, \qquad (k, n = 0, 1, 2, \cdots). \tag{9.5}$$

Proof. (9.5) is true when $k = 1$ by Lemma 9.1. Its general validity follows directly by a brief induction on k.

10. We can now establish the numerical periodicity of (h) modulo p.[21]

THEOREM 10.1. *Let (h) be an elliptic divisibility sequence, and let p be any prime which divides neither h_2 nor h_3. Let ρ be the rank of apparition of p in (h), and let τ be the least positive integer such that*

$$(-b)^{\tau^2} \equiv 1, \qquad c^{\tau} \equiv 1 \pmod p \tag{10.1}$$

when b and c are the integers specified in Theorems 8.1 and 8.2. Then (h) is purely periodic modulo *p with period $\tau\rho$.*

Proof. The proof of this theorem depends on the following lemma whose proof is left to the reader.

LEMMA 10.1. *If τ is defined as in Theorem 10.1 and if k is an integer such that*

$$(-b)^{k^2} \equiv 1, \qquad c^{k} \equiv 1 \pmod p \tag{10.2}$$

then τ divides k.

We see from (10.1) and the congruence (9.5) of Theorem 9.1 that $\tau\rho$ is a period of (h) and (h) is purely periodic modulo p. Hence by Theorem 5.2, any other period π of (h) modulo p is a multiple of ρ; say $\pi = k\rho$. But if $k\rho$ is a period, then on taking n equal to 1 and 2 in (9.5), we obtain the congruences

$$(-c)^k b^{k^2} \equiv 1, \qquad (-c)^k c^k b^{k^2} \equiv 1 \pmod p.$$

Since k and k^2 have the same parity, (10.2) follows. Hence, τ divides k, so that $\tau\rho$ divides π. This completes the proof of the theorem.

11. This section is devoted to the proof of the Theorem 11.1 quoted in Section 7 in which the integer τ was explicitly determined. We shall need the following arithmetical lemma whose proof we leave to the reader.

---

[21] Periodicity for an arbitrary modulus m is an easy consequence. See Chapter VIII.

LEMMA 11. 1. *Let p be an odd prime,*[22] *d an integer prime to it, and belonging to the exponent δ* modulo p. *Then if δ is odd, there exists no integer x such that the congruence*

$$d^x \equiv -1 \pmod{p} \tag{11.2}$$

is satisfied. But if δ is even, (11. 2) *is satisfied if and only if x is an odd multiple of δ.*

We observe first that the congruences (11. 1) and (8. 3) allow us to identify the integers c of Theorems 11. 1 and 8. 2. Since p is a prime, the congruence (8. 4) implies that b is congruent to either plus or minus one. Assume that

$$b^\tau \equiv +1 \pmod{p}. \tag{11.3}$$

Then by (10. 1), $(-b)^{\tau^2} \equiv (-1)^\tau \equiv 1 \pmod{p}$. Hence τ *must be even.* Now by (8. 3), $b^\tau \equiv h^\tau{}_{\rho-1}c^\tau$. Since $c^\tau \equiv 1$ by (10. 1), (11. 3) gives

$$h^\tau{}_{\rho-1} \equiv 1 \pmod{p}. \tag{11.4}$$

Then by (11. 1), (10. 1),

$$e^\tau \equiv 1 \pmod{p}. \tag{11.5}$$

Let $\sigma = [\epsilon, \kappa]$ be the least common multiple of the exponents to which e and $h_{\rho-1}$ belong modulo p. Then (11. 4) and (11. 5) imply that $\kappa \mid \tau$, $\epsilon \mid \tau$. Hence

$$\sigma \mid \tau. \tag{11.6}$$

On the other hand, $h^\sigma{}_{\rho-1} \equiv 1$ and $e^\sigma \equiv 1 \pmod{p}$. Hence by (11. 1) and (8. 3),

$$c^\sigma \equiv 1, \qquad b^\sigma \equiv 1 \pmod{p}. \tag{11.7}$$

Now if σ is even, (11. 7) implies that $c^\sigma \equiv 1$, $(-b)^\sigma \equiv 1 \pmod{p}$. Hence by Lemma 10. 1, $\tau \mid \sigma$, so that by (11. 6), $\tau = \sigma$.

σ is odd if and only if both ϵ and κ are odd. In this case (11. 7) implies that $c^{2\sigma} \equiv 1$, $(-b)^{4\sigma^2} \equiv 1 \pmod{p}$. Hence by Lemma 10. 1, $\tau \mid 2\sigma$. But τ is even and by (11. 6), σ divides τ. Hence $\tau = 2\sigma$. This disposes of the first case of the theorem.

Assume now that

$$b^\tau \equiv -1 \pmod{p}. \tag{11.8}$$

Then by (8. 3),

[22] The lemma is false if $p = 2$.

(11.9) $$h^{\tau}{}_{\rho-1} \equiv -1 \pmod{p},$$

and by (8.3) and (11.1)

(11.10) $$e^{\tau} \equiv -1 \pmod{p}.$$

Now by Lemma 11.1, (11.9) and (11.10) imply that both κ and ϵ are even, and that τ is both an odd multiple of $\kappa/2$ and an odd multiple of $\epsilon/2$. But if σ now denotes $[\epsilon/2, \kappa/2]$,

(11.11) $$\sigma \mid \tau.$$

Hence σ must be an odd multiple of both $\epsilon/2$ and $\kappa/2$. It follows that if (11.8) holds, both ϵ and κ must be even and both divisible by precisely the same power of two.

Assume, conversely, that the last mentioned conditions are satisfied. Then σ is an odd multiple of both $\epsilon/2$ and $\kappa/2$, so that by Lemma 11.1

$$h^{\sigma}{}_{\rho-1} \equiv -1, \qquad e^{\sigma} \equiv -1 \pmod{p}.$$

But then by (11.1), (8.3) and (8.8)

$$c^{\sigma} \equiv 1, \qquad b^{\sigma} \equiv -1 \pmod{p}.$$

Hence $(-b)^{\sigma^2} \equiv (-1)^{\sigma^2+\sigma} + 1$. Therefore by Lemma 10.1, $\tau \mid \sigma$. Hence by (11.11) $\tau = \sigma$. This completes the proof.

IV. The Representation of Elliptic Sequences by Elliptic Functions.

12. If (h) is a proper elliptic divisibilty sequence, we have seen that if either h_2 or h_3 vanishes, the general term of the sequence becomes a simple product of powers, and the arithmetical properties of the sequence are patent. Consider now a general elliptic divisibility sequence so that the first five values of (h) are integers and

(12.0) $$h_0 = 0, \qquad h_1 = 1, \qquad h_2 h_3 \neq 0; \qquad h_2 \mid h_4.$$

We shall devote this chapter to the proof of the following fundamental result.

THEOREM 12.1. *If (h) is a general elliptic divisibility sequence, there exist two rational numbers g_2 and g_3 and a complex constant u such that if $\wp(w; g_2, g_3)$ is the Weierstrass function with invariants g_2 and g_3, then*

(12.1) $$h_n = \psi_n(u) = \sigma(nu)/\sigma(u)^{n^2}.$$

Here $\sigma(w)$ is the Weierstrass sigma function.

Proof. Let (h) be a general elliptic divisibility sequence. Since $\psi_n(w)$ is always a solution of (1.1) and $\psi_0(w) = 0$, $\psi_1(w) = 1$, it suffices to show that we can determine g_2, g_3 and u so that:

(12.2) $(\alpha): \psi_2(u) = h_2$; $(\beta): \psi_3(u) = h_3$; $(\gamma): \psi_4(u) = h_4$.

We quote for reference eight familiar formulas of elliptic function theory:

(12.3) $\psi_2(w) = - \wp'(w)$.

(12.4) $\psi_3(w)\ 3\wp^4(w) - \tfrac{3}{2}g_2\wp^2(w) - 3g_3\wp(w) - \tfrac{1}{16}g_2^2$.

(12.5) $\wp(2w) - \wp(w) = \tfrac{1}{4}\left(\dfrac{\wp''(w)}{\wp'(w)}\right)^2 - 3\wp(w)$

(12.6) $\wp(3w) - \wp(w) = \wp'^2(w)\ (\wp'^4(w) - \psi_3(w)\wp''(w)) \div \psi_3^2(w)$.

(12.7) $\wp(2w) - \wp(w) = -\dfrac{\psi_1(w)\psi_3(w)}{\psi_2^2(w)}$.

(12.8) $\wp(3w) - \wp(w) = -\dfrac{\psi_2(w)\psi_4(w)}{\psi_3^2(w)}$.

(12.9) $\wp'^2(w) = 4\wp^3(w) - g_2\wp(w) - g_3$.

(12.10) $\wp''(w) = 6\wp^2(w) - g_2/2$.

From (12.10):

(12.11) $g_2 = 12\wp^2(w)) - 2\wp''(w)$.

From (12.9) and (12.10):

(12.12) $g_3 = 2\wp(w)\ (\wp''(w) - 4\wp^2(w) - \wp'^2(w))$.

13. *Proof* (Continued). Now assume that the conditions (12.2) (α), (β), (γ) can be satisfied. Then since $\psi_1(u) = 1$, (12.1), (12.3), (12.7) and (12.8) give:

(13.1) $\wp'(u) = - h_2$,

(13.2) $\wp(2u) - \wp(u) = - h_3/h_2^2$,

(13.3) $\wp(3u) - \wp(u) = - h_2h_4/h_3^2$.

Now by (12.6), (13.3) and (13.1):

$$- h_2h_4/h_3^2 = h_2^2/h_3^2(h_2^4 - h_3\wp''(u)).$$

Hence solving for $\wp''(u)$:

(13.4) $\wp''(u) = (h_2^5 + h_4)/h_2h_3.$

Next, using (13.2), (12.5) and (13.1), (13.4):

$$-h_3/h_2^2 = \tfrac{1}{4}\{(h_2^5 + h_4)/-h_2^2h_3\}^2 - 3\wp(u).$$

Hence solving for $\wp(u)$:

(13.5) $\wp(u) = (h_4^2 + 2h_2^5h_4 + 4h_2^2h_3^3 + h_2^{10}) \div 12h_2^4h_3^2.$

Next, using (12.11), (13.5) and (13.4):

(13.6) $g_2 = (h_2^{20} + 4h_2^{15}h_4 - 16h_2^{12}h_3^3 + 6h_2^{10}h_4^2 - 8h_2^7h_3^3h_4$
$+ 4h_2^5h_4^3 + 16h_2^4h_3^6 + 8h_2^2h_3^3h_4^2 + h_4^4) \div 12h_2^8h_3^4.$

Finally, using (12.12), (13.5), (13.4) and (13.6):

(13.7) $g_3 = -(h_2^{30} + 6h_2^{25}h_4 - 24h_2^{22}h_3^3 + 15h_2^{20}h_4^2 - 60h_2^{17}h_3^3h_4$
$+ 20h_2^{15}h_4^3 + 120h_2^{14}h_3^6 - 36h_2^{12}h_3^3h_4^2 + 15h_2^{10}h_4^4$
$- 48h_2^9h_3^6h_4 + 12h_2^7h_3^3h_4^3 + 64h_2^6h_3^9 + 6h_2^5h_4^5$
$+ 48h_2^4h_3^6h_4^2 + 12h_2^2h_3^3h_4^4 + h_4^6) \div 216h_2^{12}h_3^6.$

(13.5), (13.6) and (13.7) are *necessary* conditions that the equations (12.2) hold. Now since by (12.1) neither h_2 nor h_3 is zero, we can start by *defining* g_2, g_3 and u (13.6), (13.7) and (13.5). Then u is determined save for sign up to a period of $\wp(w)$.

On combining (13.5) and (13.6), we find that

$$g_2 - 12\wp^2(u) = -2(h_2^5 + h_4)/h_2h_3.$$

Hence (13.4) follows from formula (12.11).

Now combining (13.7) with (13.5), (13.4) and (13.6), we obtain the formula

$$g_3 - 2\wp(u)[\wp''(u) - 4\wp^2(u)] = -h_2^2.$$

Hence by formula (12.12), $\wp'^2(u) = h_2^2$. We now choose the sign of u so that (13.1) is satisfied. u is now fixed up to a period of the $\wp$ function. But then (12.2)α follows immediately from formula (12.3).

Next, using (12.5) and substituting in it for $\wp'(u)$, $\wp''(u)$ and $\wp(u)$ from (13.1), (13.4) and (13.5), we find that $\wp(2u) - \wp(u) = -h_3/h_2^2$.

Hence (12.2)β follows from (12.7), (12.2) and the fact that $\psi_1(u) = 1$.

Finally on substituting on the right of (12.6) for $p'(u)$, $p''(u)$ and $\psi_3(u)$, we find that $\wp(3u) - \wp(u) = -h_2h_4/h_3^2$.

Hence (12.2)γ follows from (12.8 and (12.2)α and β.

V. The Relationship Between the Numerical Periodicity of a Sequence and the Periodicity of the Corresponding Elliptic Functions.

14. We shall now prove Theorem 8.1 of Chapter III. Throughout this part of the paper, (h) denotes a fixed general elliptic sequence, and p a fixed prime greater than three dividing neither h_2 nor h_3. For convenience of printing, the rank of apparition of p in (h) will be denoted by r, rather than by ρ as heretofore.

It follows from the results of Part IV that

$$(14.1) \qquad h_n = \psi_n(u).$$

Furthermore g_2, g_3 and $\wp(u)$ are integers modulo p.

We commence by stating the results of elliptic function theory which we shall need.[23] If we regard w in $\psi_n(w)$ as a complex variable, $\psi_n(w)$ may be expressed in terms of the Weierstrass signma function as follows:

$$(14.2) \qquad \psi_n(w) = \sigma(nw)/\sigma(w)^{n^2}.$$

If 2ω is a period of the $\wp$ function, then with the usual notations of the theory of elliptic functions,

$$(14.3) \qquad \sigma(w + 2\omega) = -e^{2\eta(w+\omega)}\sigma(w).$$

On the other hand, if $z = \wp(w)$, $\psi_n(w)$ may be expressed as a polynomial in z, say $F_n(z)$, of the form

$$(14.4) \qquad \psi_n(w) = F_n(z) = e_q \sum_{r=0}^{q} A_{q-r} z^r$$

where the degree q of $F_n(z)$ in z is $(n^2 - 1)/2$ or $(n^2 - 4)/2$, and e_q is 1 or $h_2/2$ according as n is odd or even. The coefficients A of $F_n(z)$ are polynomials in $g_2/4$ and g_3 with rational integral coefficients:

$$(14.5) \qquad A_k = A_k(g_2/4, g_3), \qquad k = 0, 1, \cdots, q.$$

Hence each A_k is an integer modulo p. Furthermore A_k is homogeneous of degree k if g_2 is given the weight two and g_3 the weight three. In particular,

[23] See Fricke, *Die Elliptischen Funktionen* . . . II, Berlin, 1922, pp. 184-205.

(14.6) $\quad A_0 = n,$

(14.7) $\quad A_1 = 0, \qquad A_2 = bg_2/4, \qquad A_3 = cg_3, \qquad A_4 = dg_2^2/16$

where b, c, d are integers depending of course on n.

It is also well known that if we consider the roots ζ of $F_r(z) = 0$ (where it will be recalled that r is the rank of apparition of p in (h)) then each ζ may be expressed in the form

(14.8) $$\zeta = \wp\,(2\omega/r)$$

where 2ω is some period of the p function.

15. Let $\mathfrak{R}$ denote the field obtained by adjoining all the roots of $F_r(z) = 0$ to the field of rationals, and let $\mathfrak{p}$ denote any prime ideal divisor of p in $\mathfrak{R}$. By Theorem 5.1, the rank of apparition r of p is less than $2p + 2$. Hence either r is prime to p, or $r = p$, or $r = 2p$.

We shall assume that r is prime to p in this section. It follows from the results on $F_n(z)$ stated in the previous section, that all the roots ζ of $F_r(z) = 0$ are algebraic integers modulo p and that we have the congruence

$$h_r \equiv c_r \prod_{(\zeta)} (\wp\,(u) - \zeta) \pmod{p}$$

where c_r is an integer prime to p. But by the definition of r, h_r is divisible by p. Hence we have the congruence in $\mathfrak{R}$: $\Pi(\wp\,(u) - \zeta) \equiv 0 \pmod{\mathfrak{p}}$. Since $\mathfrak{p}$ is a prime ideal, there must exist by (14.8) a period 2ω of the $\wp$ function such that

(15.1) $$\wp\,(u) \equiv \wp\,(2\omega/r) \pmod{\mathfrak{p}}.$$

We deduce from (14.4) and (14.1) that

(15.2) $$h_n \equiv \psi_n(2\omega/r) \pmod{\mathfrak{p}}$$

for $n = 0, 1, 2, \cdots$.

Consider now $\psi_{r-n}(2\omega/r)$ where $0 \leqq n \leqq r$. By formulas (14.2) and (14.3);

$$\begin{aligned}\psi_{r-n}(2\omega/r) &= \sigma(-2n\omega/r + 2\omega) \div \sigma(2\omega/r)^{r^2-2rn+n^2} \\ &= \alpha^n\beta\sigma(2n\omega/r) \div \sigma(2\omega/r)^n = \alpha^n\beta\psi_n(2\omega/r).\end{aligned}$$

Here $\alpha = e^{4\eta\omega/r}\sigma(2\omega/r)^{2r}$, and $\beta = e^{2\eta\omega} \div \sigma(2\omega/r)^{r^2}$, and we have used the fact that $\sigma(-w) = -\sigma(w)$. (15.2) now gives the congruences

(15.3) $$h_{r-n} \equiv \alpha^n\beta h_n \pmod{\mathfrak{p}}.$$

Letting n equal one and two in (15.3), we see that $\alpha\beta$ and $\alpha^2\beta$ are congruent to rational integers modulo $\mathfrak{p}$. Hence α and β are congruent modulo $\mathfrak{p}$ to two rational integers; say a and b. Thus (15.3) becomes

$$h_{r-n} \equiv a^n b h_n \pmod{\mathfrak{p}}.$$

Since all the Roman letters denote rational integers, we deduce that

$$h_{r-n} \equiv a^n b h_n \pmod{p}.$$

On replacing r by ρ, we obtain Theorem 8.1 for the case when the rank of apparition of the prime p is not p or $2p$.

16. It remains to discuss the more difficult case, when the rank of appartition r of p equals p or $2p$. It follows from the form of the coefficients A_k of $F_r(z)$, that if p divides both g_2 and g_3, it divides every coefficient of $F_r(z)$. The converse is also true.

LEMMA 16.1. *A necessary and sufficient condition that p divide every coefficient of $F_p(z)$ or $F_{2p}(z)$ is that p divide both g_2 and g_3. The rank of apparition of every such prime is p.*

Proof. We need only prove the necessity of the condition. Assume that $r = p$ and

$$A_k \equiv 0 \pmod{p}, \qquad k = 0, 1, \cdots, q = (p^2 - 1)/2.$$

Let $\mathfrak{G}$ denote the Galois field obtained by adjoining to the field of rationals the three roots e, e_2, e_3 of $4x^3 - g_2 x - g_3 = 0$. Then with the usual notation, $e_i = \wp(\omega_i)$, $(i = 1, 2, 3)$ where $2\omega_i$ is a period and $\omega_1 + \omega_2 + \omega_3 = 0$. The numbers e_i are integers modulo p since p is odd. Now let $\mathfrak{p}$ be any prime ideal divisor of p in $\mathfrak{G}$. Then by (14.2), (14.4) and our hypothesis on the A_h,

$$(16.1) \qquad \sigma(p\omega_i)/\sigma(\omega_i)^{p^2} = \psi_p(\omega_i) = e_q \sum_{r=0}^{q} A_{q-r} e_i^{\,r} \equiv 0 \pmod{p}.$$

On the other hand on writing $p = (2p-1)/2 + 1$ and using the periodicity of the sigma function,

$$\sigma(p\omega_i) = (-1)^{(p-1)/2} e^{2\eta(p-1)/2(\omega_i + [(p-1)/2]\omega_i)} \sigma(\omega_i).$$

But

$$e^{\frac{1}{2}\eta\omega_i} = (e_i - e_j)^{1/4}(e_i - e_k)^{1/4}\sigma(\omega_i).$$

Hence

$$\sigma(p\omega_i) = (e_i - e_j)^{(p^2-1)/4}(e_i - e_k)^{(p^2-1)/4}\sigma(\omega_i)^{p^2}.$$

so that

$$\psi_p(\omega_i) = (e_i - e_j)^{(p^2-1)/4} (e_i - e_k)^{(p^2-1)/4}.$$

Hence by (16.1),

$$(e_i - e_j)^{(p^2-1)/4} (e_i - e_k)^{(p^2-1)/4} \equiv 0 \pmod{\mathfrak{p}}, \quad i, j, k = 1, 2, 3, i \neq j, i \neq k.$$

Since $\mathfrak{p}$ is a prime ideal, we deduce that $e_1 \equiv e_2 \equiv e_3 \pmod{\mathfrak{p}}$. But then for every integer ξ of $\mathfrak{F}$, $4\xi^3 - g_2\xi - g_3 \equiv 4(\xi - e_1)^3 \pmod{\mathfrak{p}}$. Hence

$$e_1 \equiv e_2 \equiv e_3 \pmod{\mathfrak{p}} \text{ so that } g_2 \equiv g_3 \equiv 0 \pmod{\mathfrak{p}}.$$

Since g_2 and g_3 are rational integers modulo p, it follows that $g_2 \equiv g_3 \equiv 0 \pmod{p}$. This completes the proof of the lemma for the case when $r = p$. The proof for the case when all the coefficients of $\psi_{2p}(w)$ are divisible by p is similar and will be omitted here.

17. In view of Lemma 16.1, we need consider only the case

$$(17.1) \qquad h_r \equiv 0 \pmod{p}, \qquad r = p \text{ or } r = 2p;$$

(17.2) g_2 and g_3 not both divisible by p.

We first develop some simple arithmetical concepts which are needed in the proofs that follow. Let $\mathfrak{p}$ be a prime ideal of an algebraic number field, and α any field element. Then the principal ideal $[\alpha]$ has a unique representation of the form $[\alpha] = \mathfrak{p}^{-a}\mathfrak{b}\mathfrak{c}^{-1}$ where $\mathfrak{b}$ and $\mathfrak{c}$ are integral ideals which are co-prime and also prime to $\mathfrak{p}$, and the exponent a is a rational integer. We call a "the index of α (modulo p)." α is said to be integral modulo p if and only if its index is negative or zero, and fractional modulo p if and only if its index is positive.

The following lemmas follow readily, and their proofs are left to the reader.

LEMMA 17.1. *If α is a fraction modulo p and β is an integer modulo $\mathfrak{p}$, $\alpha \pm \beta$ is a fraction with the same index as α, and the index of $\alpha\beta$ is not greater than that of α.*

LEMMA 17.2. *If $\alpha_1, \alpha_2, \cdots, \alpha_k$ are fractions modulo $\mathfrak{p}$, the index of their product is the sum of the indices of the separate factors.*

LEMMA 17.3. *If $\alpha_1, \alpha_2, \cdots, \alpha_k$ are fractions modulo $\mathfrak{p}$, the index of $(\phi - \alpha_1)(\phi - \alpha_2) \cdots (\phi - \alpha_k)$ is the same for all ϕ which are integers modulo $\mathfrak{p}$, and equals the sum of the indices of $\alpha_1, \alpha_2, \cdots, \alpha_k$.*

18. We may now complete the proof of Theorem 8.1 as follows. With the notation of Section **16**, let the roots of $F_r(z) = 0$ be $\zeta_1, \zeta_2, \cdots, \zeta_q$. The leading coefficient of $F_r(z)$ is divisible by p but not by p^2 by formulas (14.4) and (14.6). Furthermore, there exists at least one coefficient A_k which is not divisible by p. Consequently, if $\mathfrak{p}$ as before denotes a prime ideal divisor of p in the field $\mathfrak{R}$, not all the roots ζ are integers modulo $\mathfrak{p}$. We shall now prove

LEMMA 18.1. *Not all the roots ζ of $F_r(z) = 0$ are fractions modulo $\mathfrak{p}$.*

Proof. Let ϕ denote a variable whose range is the set of all field elements of $\mathfrak{R}$ which are integers modulo $\mathfrak{p}$. Then if all the ζ are fractions, the index of $(\phi - \zeta_1)(\phi - \zeta_2), \cdots (\phi - \zeta_q)$ by Lemma 17.3 is a positive number independent of the choice of ϕ. But by formula (13.5), $z = \wp(u)$ is an admissible value of ϕ, since $\mathfrak{p}$ is prime to $6h_2h_3$. But by formulas (14.1) and (14.4),

$$h_r = F_r(z) = pl(z - \zeta_1)(z - \zeta_2) \cdots (z - \zeta_q) \tag{18.1}$$

where l is an integer depending on r but prime to p.

Now suppose that the highest power of $\mathfrak{p}$ dividing p is the k-th. Then by (17.1), the index of the left side of (18.1) is at most $-k$. But by Lemma 17.2, the index of the right side of (18.1) is greater than $-k$. This contradiction establishes the lemma.

Now let $\zeta_1, \zeta_2, \cdots, \zeta_s$ be the roots of $F_r(z) = 0$ which are integers modulo $\mathfrak{p}$, and $\zeta_{s+1}, \zeta_{s+2}, \cdots, \zeta_q$ be the roots which are fractions modulo $\mathfrak{p}$. In view of what we have just proved, both these sets of roots are non-empty. Now re-write (18.1) as

$$h_r = pl[(z - \zeta_1) \cdots (z - \zeta_s)][(z - \zeta_{s+1}) \cdots (z - \zeta_q)].$$

The index of the right side is at most equal to the index $-k$ of p. But the index of $[(z - \zeta_{s+1}) \cdots (z - \zeta_q)]$ is positive. Consequently the index of $[(z - \zeta_1) \cdots (z - \zeta_s)]$ must be negative. But this implies that $(z - \zeta_1) \cdots (z - \zeta_s) \equiv 0 \pmod{\mathfrak{p}}$. Since $\mathfrak{p}$ is a prime ideal and each term $z - \zeta_1, \cdots z - \zeta_s$ is an integer modulo $\mathfrak{p}$, there exists a ζ such that $\wp(u) = z \equiv \zeta \pmod{\mathfrak{p}}$. Hence we obtain again from (14.8) the congruence (15.1) for a suitably chosen period 2ω of the $\wp$-function. The remainder of the proof now follows exactly as in Section **15** for the case r prime to p.

VI. Equivalent Sequences. Singular Sequences and Their Representations by Circular Functions.

19. Two sequences (u) and (v) (which need neither be integral, nor solutions of (1.1)) are said to be "equivalent" if and only if there exists a constant $c \neq 0$ such that

$$u_n = c^{n^2-1} v_n, \qquad (n = 0, 1, 2, \cdots).$$

We write $(u) \sim (v)$ or $(u) = c(v)$ if it is desired to bring the constant c explicitly in evidence. $\sim$ is evidently an equivalence relation in the technical sense. We shall show in Chapter VII that there are only four types of non-equivalent solutions of (1.1), of which the elliptic function and circular function solutions are the two most important. We shall continue the further development of the properties of equivalence in section twenty-one of this chapter.

Let (h) be a general elliptic sequence. We have seen in Chapter IV that there then exists an elliptic function $\wp(w) = \wp(w; g_2, g_3)$ whose invariants g_2 and g_3 are certain rational functions of the initial values of (h), such that for a properly chosen value u of the complex variable w,

$$h_n = \sigma(nu)/\sigma(u)^{n^2}. \tag{19.1}$$

By the "discriminant" of the sequence (h) we mean the discriminant of the corresponding $\wp$-function:

$$\Delta = g_2^3 - 27 g_3^2. \tag{19.2}$$

We write $\Delta = \Delta(h)$, or $\Delta = \Delta(h_2, h_3, h_4)$ if we wish to emphasize the dependence of Δ on the initial values of (h).

If we substitute for g_2 and g_3 in (19.2) their expressions in terms of h_2, h_3 and h_4 given by formulas (13.6) and (13.7), we find that

$$\begin{aligned}\Delta(h_2, h_3, h_4) = 1/h_2^8 h_3^3 \{ & h_4^4 + 3h_2^5 h_4^3 + (3h_2^8 + 8h_3^3) h_4^2 \\ & + h_2^7 (h_2^8 - 20 h_3^3) h_4 + h_2^4 h_3^3 (16 h_3^3 - h_2^8) \}.\end{aligned} \tag{19.3}$$

The sequence (h) is said to be "singular" if and only if its discriminant $\Delta(h)$ vanishes. We shall show that a sequence is singular if and only if it is essentially a Lucas function. The main step in the proof of this result is the following theorem:

THEOREM 19.1. *Necessary and sufficient conditions that a general elliptic*

sequence (h) *be singular are that there exist integers* r *and* s *such that* $rs(r^2-s^2)\neq 0$ *and*

$$(19.4)\qquad h_2=r,\qquad h_3=s(r^2-s^3),\qquad h_4=rs^3(r^2-2s^3).$$

20. This section is devoted to the proof of Theorem 19.1. We first prove that the conditions (19.4) are ncessary for (h) to be singular. Assume then that (h) is a general elliptic sequence for which $\Delta(h)$ vanishes.

Since h_2 and h_3 are not zero and h_2 divides h_4, it follows from (19.3) that if we let

$$(20.1)\qquad u=h_2^4,\qquad v=h_3^3,\qquad w=h_4/h_2,$$

then Δ vanishes if and only if the diophantine equation

$$(20.2)\qquad 16v^2-(u^2+20uw-8w^2)v+w(u+w)^3=0$$

has solutions of the form (20.1); that is, u a perfect fourth power and v a perfect cube.

If we solve (20.2) for v by the quadratic formula, we find that

$$(20.3)\qquad 32v=u^2+20uw-8w^2\pm\sqrt{u(u-8w)^3}.$$

Hence it is necessary that $u(u-8w)$ be a square. But u is a square by (20.1). Hence we may write

$$(20.4)\qquad u=l^2=h_2^4,$$

$$(20.5)\qquad u-8w=m^2=h_2^4-8h_4/h_2,$$

where l and m are integers. Then

$$(20.6)\qquad w=(l^2-m^2)/8.$$

We find from (20.4) and (20.6) that

$$u^2+2uw-8w^2=\tfrac{1}{8}(27l^4-18l^2m^2-m^4),$$

$$\sqrt{u(u-8w)^3}=lm^3.$$

On substituting these expressions into (20.3) and multiplying by eight, we find that $256v=27l^4-18l^2m^2\pm 8lm^3-m^4$.

The right hand side of this expression factors into $(l\pm m)(3l\mp m)^3$. Hence on multiplying by two and substituting h_3^3 for v, we obtain the formula

$$(20.7)\qquad (8h_3)^3=(2l\pm 2m)(3l\mp m)^3.$$

Hence $2l \pm 2m$ is the cube of an even integer, and we may write $2l \pm 2m = (2s)^3$ where s is an integer. Now $3l \mp m + 4s^3 = 4l$. Hence $3l \mp m$ in (20.7) is divisible by four. We thus have for integral s and q

$$(20.8) \qquad l \pm m = 4s^3, \qquad 3l \mp m = 4q,$$

and (20.7) becomes $(8h_3)^3 = (8sq)^3$. Hence

$$(20.9) \qquad h_3 = sq$$

and on solving (20.8) for l and m, we find that

$$(20.10) \qquad l = s^3 + q, \qquad m = \pm\,(3s^3 - q).$$

On substituting these expressions for l and m into (20.6), we find that

$$(20.11) \qquad h_4/h_2 = w = s^3(q - s^3).$$

Finally, (20.4) and (26.10) give

$$(20.12) \qquad h_2{}^2 = s^3 + q.$$

Now let $h_2 = r$. Then by (20.12), $q = r^2 - s^3$. Then on substituting this expression for q into (20.9) and (20.11), we obtain the formulas (19.4). The accessory condition $rs(r^2 - s^3) \neq 0$ is needed to insure that (h) is general. The necessity of the conditions (19.4) is thus established.

The sufficiency of the conditions (19.4) is evident on retracing the steps of the proof of their necessity in reverse order. The sufficiency also follows directly by substituting into the formula (19.3) for $\Delta(h_2, h_3, h_4)$ the expressions for h_2, h_3 and h_4 in terms of r and s. The result vanishes identically in r and s.

The following theorem may be proved by elementary algebra on substituting into the formulas (13.6) and (13.7) giving g_2 and g_3 the expressions for h_2, h_3 and h_4 given in (19.4).

THEOREM 20.1. *If* (h) *is a singular elliptic sequence, then with the notation of Theorem* 19.1,

$$(20.13) \qquad g_2 = 3\{(r^2 - 4s^3)/6s^2\}^2, \qquad g_3 = -\{(r^2 - 4s^3)/6s^2\}^3.$$

Now if e, e_2 and e_3 denote as usual the roots of

$$(20.14) \qquad 4z^3 - g_2 z - g_3 = 0,$$

then $\Delta = 0$ if and only if two or more of the roots e_i are equal. Suppose that

(20.15) $$\Delta = 0, \qquad e_1 = e_2.$$

Then

(20.16) $$e_3 = -2e_1$$

and

(20.17) $$g_2 = 3e_3^2, \qquad g_3 = e_3^3.$$

Hence we obtain the following corollary to Theorem 20.1:

THEOREM 20.2. *If* (h) *is a singular sequence, then the roots of* (20.14) *are*

$$-(r^2 - 4s^3)/6s^2, \qquad (r^3 - 4s^3)/12s^2, \qquad (r^2 - 4s^3)/12s^2.$$

Furthermore

(20.18) $$g_2 = g_3 = 0 \qquad \textit{if and only if } r^2 = 4s^3.$$

In this case, $e_1 = e_2 = e_3 = 0$.

21. We shall now resume our discussion of the notion of equivalence of sequences introduced at the beginning of this chapter.

A sequence (α) of algebraic numbers is said to be "essentially integral" if it is equivalent to an integral sequence; that is, if there exists an algebraic number β other than zero such that $\beta^{n^2-1}\alpha_n$ is a rational integer for every n.

THEOREM 21.1. *If a sequence* (u) *is a particular solution of* (1.1), *so are all sequences equivalent to it. Furthermore, if* (u) *is general, so are all its equivalent sequences.*

THEOREM 21.2. *If an elliptic sequence* (h) *admits an elliptic function representation by means of* $\wp(w) = \wp(w; g_2, g_3)$ *and* $(k) = c(h)$ *is any equivalent sequence, then* (k) *admits an elliptic function representation by means of* $\wp(w') = \wp(w'; g'_2, g'_3)$ *where* $w' = w/c$, $g'_2 = c^4 g_2$, $g'_3 = c^6 g_3$.

Equivalence is thus the analogue of the $\wp$-function homogeneity relation: $\wp(w/c; c^4 g_2, c^6 g_3) = c^2 \wp(w; g_2, g_3)$.

The proofs of these two theorems are almost immediate and are left to the reader.

THEOREM 21.3. *Every proper solution of* (1.1) *in rational numbers is essentially integral, and equivalent to an integral divisibility-sequence.*

Proof. Let (a) be a proper rational solution of (1.1) so that $a_0 = 0$, $a_1 = 1$ not both a_2, a_3 vanish and a_n is rational. If a_2 is zero, the theorem is

obvious from Lemma 4.1 of Chapter II, formula (4.13); for we may take c^8 equal to the denominator of a_3. If a_2 is not zero, (a) is clearly uniquely determined by the initial values of a_2, a_3 and a_4. Now we may write $a_2 = c_2/a$, $a_3 = c_3/a$, $a_4 = c_4/a$ where c_2, c_3, c_4 and a are integers and $c_2 \neq 0$. Then by Theorem 4.1, (b) is an equivalent integral divisibility sequence if $b_n = (c_2 a)^{n^2-1} a_n$.

We may now prove a converse to Theorem 12.1 of Chapter IV.

THEOREM 21.4. *If the invariants g_2 and g_3 of the function $\wp(w)$ are rational numbers and if u is such that $\wp(u)$ is rational, then $a_n = \psi_n(u)$ is equivalent to an integral elliptic divisibility sequence.*

Proof. By (14.4), all the a_n are rational. But $\psi_n(w)$ satisfies (1.1). Hence the result follows from the previous theorem.

22. We shall resume the development of the properties of singular solutions by establishing the following theorem:

THEOREM 22.1. *Every singular elliptic sequence is equivalent either to the sequence*

$$0, 1, 2, \cdots, n, \cdots \tag{22.1}$$

of the positive integers or to a Lucas sequence

$$U_0, U_1, U_2, \cdots, U_n, \cdots \tag{22.2}$$

where $U_n = (a^n - b^n)/(a - b)$, $Q = ab = 1$, and $P = a + b$ is in general, a quadratic irrationality.

Evidently such a Lucas sequence may be written in the form $U_n = \sin n\theta / \sin\theta$ for a suitably chosen complex number θ, and is hence parameterized by circular functions.

Proof. Let (h) be a singular elliptic divisibility sequence so that $\Delta(h) = 0$. Suppose first that $g_2 = g_3 = 0$. Then it follows from Theorem 20.2 that

$$r^2 = 4s^3 \tag{22.3}$$

where r and s are the integers introduced in Theorem 19.1. But the diophantine relation (22.3) implies that there exists an integer c such that $r = 2c^3$, $s = c^2$. Then by (19.4), $h_2 = c^3 2$, $h_3 = c^8 3$, $h_4 = c^{15} 4$.

Hence by Theorem 4.1, (h) is equivalent to the solution (22.1).

Now assume that not both g_2 and g_3 are zero. We first develop some lemmas.

LEMMA 22.1. *Let α and β be two distinct numbers neither of which is zero, and let $p = \alpha + \beta$, $q = \alpha\beta$, $u_n = (\alpha^n - \beta^n)/(\alpha - \beta)$. Then*

$$q^{(1-n)/2}u_n \tag{22.4}$$

is a solution of (1.1).

This lemma, as was mentioned in the introduction, is due to Lucas. We call (22.4) a "Lucas solution" of (1.1), or a "Lucas sequence."

In Lucas' arithmetical theory, p and q are rational integers so that a Lucas solution is not generally an integral sequence, although it is evidently equivalent to an integral sequence.

We may however restate the lemma of Lucas in a way which overcomes this defect and is more convenient for our purposes. Since neither α nor β is zero, we may let $a = \sqrt{\alpha/\beta}$ and $b = \sqrt{\beta/\alpha}$. Then $P = a + b = p/\sqrt{q}$ and $Q = ab = 1$ while $U_n = (a^n - b^n)/(a - b) = q^{(1-n)/2}u_n$. Hence we may state the following modification of Lemma 22.1.

LEMMA 22.2. *Every Lucas solution of* (1.1) *is of the form*

$$U_n = (a^n - b^n)/(a - b) \tag{22.5}$$

where $P = a + b$ and $Q = ab = 1$.

We shall assume that[24] $P \neq 0$, as otherwise (U) is not general.

The initial values of the Lucas solution (22.5) with $Q = 1$ are

$$0, 1, P, P^2 - 1, P^3 - 2P.$$

On comparing these values with (19.4), we obtain the following result:

LEMMA 22.3. *A necessary and sufficient condition that a general*[25] *elliptic sequence be a Lucas solution of* (1.1) *is that it be a singular solution with $r = P$ and $s = 1$.*

Now consider any singular sequence (h) with g_2 and g_3 not both zero. By Theorem 19.1,

$$h_2 = r, \qquad h_3 = s(r^2 - s^3), \qquad h_4 = rs^3(r^2 - 2s^3) \tag{19.4}$$

where r and s are integers and $rs(r^2 - s^3) \neq 0$.

---

[24] Lucas solutions of (1.1) with $P = 0$ are discussed in Chapter VII.

[25] Solutions with $h_2 = 0$, $h_3 \neq 0$ are equivalent to a Lucas solution. Solutions with $h_2 \neq 0$, $h_3 = 0$ are not generally Lucas solutions. See Chapter VII.

Now let $r = c^3P$, $s = c^2$. Then c is in general a quadratic irrationality. Hence P is in general a quadratic irrationality; namely

$$P = r\sqrt{s}/s^2. \tag{22.6}$$

Then (19.4) becomes

$$h_2 = c^3P, \qquad h_3 = c^8(P^2 - 1), \qquad h_4 = c^{15}(P^2 - 2).$$

Hence (h) is equivalent to a Lucas solution with P given by (22.6) and $Q = 1$. This completes the proof of Theorem 22.1.

VII. Special Sequences.

23. We have seen that any sequence (h) whose initial values satisfy the conditions

$$h_0 = 0, \qquad h_1 = 1, \qquad h_2 \neq 0, \qquad h_3 \neq 0 \tag{23.1}$$

may be parameterized by elliptic or circular functions. We discuss now the special sequences which arise when one or more of the conditions (23.1) are violated. Until further notice, (h) denotes a sequence of complex numbers satisfying (1.1), so that

$$h_{m+n}h_{m-n} = h_{m+1}h_{m-1}h_n^2 - h_{n+1}h_{n-1}h_m^2, \qquad m \geqq n \geqq 1. \tag{4.11}$$

Sequences in which $h_1^2 \neq 1$ are uninteresting. For on first letting $m = 2$ and $n = 2$ in (4.11) and then letting $m = n$, $n = 1$; $n = n$, we obtain the relations

$$h_0h_2 = 0, \qquad (h_1^2 - 1)h_{n+1}h_{n-1} = 0, \qquad n \geqq 1; \tag{23.2}$$

$$h_0h_{2n} = 0. \tag{23.3}$$

Now if $h_1^2 \neq 1$, then $h_{n+1}h_{n-1} = 0$. Hence since n is arbitrary, $h_{m+n}h_{m-n} = 0$ for $m \geqq n \geqq 1$. Since the integers $m + n$ and $m - n$ are of the same parity, there can be at most two non-vanishing terms in (h) and their suffices must be of opposite parity.

It is evident conversely that if k and l are any integers $\geqq 0$, then $h_n = 0$, $n \neq k$, $n \neq k + 2l + 1$; h_k, h_{k+2l+1} arbitrary, defines a solution of (1.1).

We shall assume henceforth that $h_1^2 = 1$. There is no loss of generality in assuming then that $h_1 = 1$; for if h_n is a solution of (1.1), so is $(-1)^n h_n$.

We consider next solutions with $h_2 = 0$. We see from (23.2) that a sufficient condition that $h_2 = 0$ is that $h_0 \neq 0$. The simplest example of such

a solution is the Lucas sequence $h_n = \sin n\pi/2$, $n > 0$. This solution is periodic with period four and purely periodic if and only if $h_0 = 0$. The Kronecker symbol solution $(-8/n)$, mentioned in the introduction, equals $(-1)^{(n^2-1)/8} \sin n\pi/2$, and is hence essentially a Lucas solution, but of period eight instead of four. Evidently the fourth term of this solution is not zero. We shall now show that there is essentially no other such solution.

THEOREM 23.1. *Every solution (h) of (1.1) with $h_1 = 1$, $h_2 = 0$ and $h_3 \neq 0$ is equivalent to the Kronecker symbol solution, and is hence a Lucas solution; that is for $n > 0$,*

$$(23.4) \qquad h_n = \begin{matrix} 0 & n \text{ even} \\ (-1)^{[n/4]} h_3^{(n^2-1)/8} & n \text{ odd.} \end{matrix}$$

Proof. It is easily verified that $[(2n+1)/4] \equiv n + 1 + n(n+1)/2 \pmod 2$. Hence an equivalent way of stating (23.4) is:

$$(23.5) \qquad h_{2n} = 0, \quad h_{2n+1} = (-1)^{n+1}(-h_3)^{n(n+1)/2}, \quad (n = 1, 2, 3, \cdots).$$

Now since (h) satisfies (4.11), we obtain on taking first $m = 2n$ and $n = 2$ and then $m = 2n-2$ and $n = 3$ the two relations

$$(23.6) \qquad h_{2n+2}h_{2n-2} = -h_1h_3h_{2n}, \qquad n \geqq 1.$$

$$(23.7) \qquad h_{2n+1}h_{2n-5} = h_{2n-1}h_{2n-3}h_3^2, \qquad n \geqq 3.$$

Since h_2 vanishes, the first part of (23.5) follows by a brief induction from (23.6). Since $h_1 = 1$, we can calculate h_{2n+1} for $n = 2$ and $n = 3$ from (4.5): $h_{2n+1} = h_{n+2}h_n^3 - h_{n-1}h^3_{n+1}$.

We thus find that $h_5 = h_3^3$, $h_7 = -h_3^6$, so that (23.5) is true for $n \leqq 3$. Its general validity now follows readily by an induction based on (23.7).

24. We next discuss solutions with vanishing fourth term. We see from (23.3) that if $h_0 \neq 0$, all terms of positive even suffix vanish. But since we are assuming that $h_1 = 1$, it follows by a brief induction based on (4.5) that all terms of odd suffix vanish save h_1. Conversely

$$(24.1) \qquad h_n = 0, \qquad n > 1$$

is evidently a solution of (1.1) regardless of the values of h_0 and h_1. We shall therefore assume henceforth that

$$(24.2) \qquad h_0 = 0, \quad h_1 = 1, \quad h_3 = 0.$$

There exist an infinite number of essentially distinct solutions of (1.1) meeting these conditions. For let l denote a fixed odd number greater than one and define a numerical function of n and l, $\lambda_n = \lambda_n(l)$ as follows:

$$(24.3)\qquad \lambda_n = \begin{matrix} 0 & \text{if} & n \not\equiv \pm 1 \pmod{l}; \\ 1 & \text{if} & n \equiv +1 \pmod{l}; \\ -1 & \text{if} & n \equiv -1 \pmod{l}. \end{matrix}$$

THEOREM 24.1. *If c is a constant and not zero then*

$$(24.4)\qquad l_n = \lambda_n c^{1-n\lambda_n}$$

is a solution of (1.1) *whose initial values satisfy the conditions of* (24.2).

In particular, on taking $c = 1$, we see that λ_n itself satisfies (1.1).[26] If $l = 3$, λ_n reduces to the Legendre symbol solution $(n/3)$ mentioned in the introduction. We see incidentally that (1.1) has integral periodic solutions with any preassigned odd period l; but such a solution is a divisibility sequence only if $l = 3$.

Proof. The initial values given by formula (24.4) are evidently $l_0 = 0$, $l_1 = 1$ and $l_3 = 0$, so that (24.2) is satisfied. If we substitute l_n into the basic recurrence (4.11), the left hand side vanishes unless $m + n \equiv \pm 1 \pmod{l}$ and $m - n \equiv \pm 1 \pmod{l}$. Hence, since l is odd, there are only four cases when the left side of (4.11) is not zero; namely[27] (i) $m \equiv 1$, $n \equiv 0$; (ii) $m \equiv 0, n \equiv 1$; (iii) $m \equiv 0$, $n \equiv -1$; (iv) $m \equiv -1, n \equiv 0$.

Now of the two terms on the right hand side of (4.11), $l_{m+1}l_{m-1}l_n^2$ vanishes unless $m \equiv 0$ and $n \equiv \pm 1$, and $l_{n+1}l_{n-1}l_m^2$ vanishes unless $n \equiv 0$ and $m \equiv \pm 1$. Hence (4.11) is satisfied except, perhaps, in the four cases just listed. The following table lists the values of $\lambda_m, \cdots$ for each of the four cases and the computed values of the terms of (4.11) which result. A glance at these completes the proof.

TABLE OF VALUES OF (l)

Case	λ_m	λ_n	λ_{m+1}	λ_{m-1}	λ_{n+1}	λ_{n-1}	λ_{m+n}	λ_{m-n}	$l_{m+n}l_{m-n}$	$l_{m+1}l_{m-1}l_n^2$	$-l_{n-1}l_{n+1}l_m^2$
(i)	1	0	0	0	1	-1	1	1	c^{2-2m}	0	c^{2-2m}
(ii)	0	1	1	-1	0	0	1	-1	$-c^{2-2n}$	$-c^{2-2n}$	0
(iii)	0	-1	1	-1	0	0	-1	1	$-c^{2-2n}$	$-c^{2-2n}$	0
(iv)	-1	0	0	0	1	-1	-1	-1	c^{2+2m}	0	c^{2+2m}.

---

[26] λ_n satisfies (1.1) if $l = 4$, but Theorem 24.1 is untrue in this case for c's chosen arbitrarily.

[27] We suppress the modulus l when no confusion can arise.

25. We shall next show that the solutions (l) just investigated are essentially the only type of solution of (1.1) with fourth term zero.

THEOREM 25.1. *Every solution (h) of (1.1) with $h_0 = 0$, $h_1 = 1$ and $h_3 = 0$ either has all its other terms zero with at most one exception, or it is equivalent to a solution (l) of the type described in Theorem 24.1.*

Proof. Let (h) be a solution of (1.1) satisfying the conditions $h_0 = 0$, $h_1 = 1$ and $h_3 = 0$. Then (4.5) holds:

$$h_{2k+1} = h_{k+2}h_k^3 - h_{k-1}h_{k+1}^3, \qquad k \geqq 1. \tag{4.5}$$

If all terms of (h) with even suffices vanish, we find by a brief induction based on (4.5) that all terms of (h) of odd suffix vanish save h_1, and we have the trivial solution (24.1) again.

If not all terms of even suffix vanish, there is a first term which does not vanish. Consequently, there exists an odd integer l not less than three such that

$$h_0 = h_2 = \cdots = h_{l-3} = 0; \tag{25.1}$$

$$h_{l-1} \neq 0. \tag{25.2}$$

I say that

$$h_l = 0 \tag{25.3}$$

and

$$h_n = 0 \quad \text{for} \quad 1 < n < l - 1 \quad \text{if} \quad l > 3. \tag{25.4}$$

For (25.3) is true by hypothesis if $l = 3$. If $l > 4$, then (25.4) is true for even n by (25.1). Hence if (25.4) were false, there would exist an integer $k > 1$ such that $h_n = 0$ for $1 < n < 2k + 1 < l$ but $h_{2k+1} \neq 0$. However, by (4.5) $h_{2k+1} = 0$, since $1 \leqq k - 1 < k + 2 < 2k + 1$. This contradiction establishes (25.4). (25.3) now follows from (25.4) on taking n equal to $(l-1)/2$ in (4.5).

It may happen that $h_{l+1} = 0$. If so, it may be readily proved by induction that $h_n = 0$ for $n > l + 1$.[28] It is evident that conversely, $h_0 = 0$, $h_1 = 1$, $h_n = 0$, $n \neq l - 1$ gives a solution of (1.1). The first part of the theorem is thus established, and we may assume for the remainder of the proof that

$$h_{l+1} \neq 0. \tag{25.5}$$

[28] For odd n, we use (4.5) as a basis for the induction. For even n, we use the formula $h_{2k}h_{l-1} = h_{m+1}h_{m-1}h_n^2 - h_{n+1}h_{n-1}h_m^2$ obtained by letting $m = k + (l-1)/2$ and $n = k - (l-1)/2$ in (4.11).

I say that

$$(25.6)\qquad h_n = 0 \quad \text{for} \quad l+1 < n < 2l-1 \quad \text{if} \quad l > 3;$$

$$(25.7)\qquad h_{2l-1} = h_{l-1}h^3_{l+1}; \quad h_{2l} = 0; \quad h_{l+1} = -h_{l-1}h^3_{l+1}.$$

For if n is even, then by (4.11)

$$(25.8)\quad h_nh_{l-1} = h_{(n+l+1)/2}h_{(n+l-3)/2}h^2_{(n-l+1)/2} - h_{(n-l+3)/2}h_{(n-l-1)/2}h^2_{(n+l-1)/2}.$$

Now if $n = l+3$, $h_{(n-l+3)/3} = h_3 = 0$ and $h_{(n-l-1)/2} = h_2 = 0$. If $n > l+3$, then $1 < (n-l-1)/2 < (n-l+1)/2 < l/2 < l-1$. Hence $h_{(n-l\pm1)/2} = 0$ by (25.4). Hence $h_n = 0$ by (25.5). If n is odd, say $n = 2k+1$, $h_n = 0$ directly by (4.5) and (25.4). The first and third equations of (23.7) follow directly from (4.5), and the second equation follows from (25.8) on putting n equal to $2l$.

We can now prove that

$$(25.9)\qquad h_n = a^{n^2-1}\lambda_n c^{1-n\lambda_n}$$

where

$$(25.10)\qquad a = (-h_{l-1}h_{l+1})^{1/2l^2},\ c = (-h_{l-1})^{(2+l)/2l^2}h_{l+1}{}^{(2-l)/2l^2}.$$

Since $\lambda_n c^{1-n\lambda_n}$ is a special (l) solution, this step will complete the proof of the theorem.

If n is less than $2l+2$ and not congruent to ± 1 modulo l, (25.9) gives $h_n = 0$ in agreement with (25.4) and (25.6). It is readily seen that (25.9) also gives the values for $h_{l\pm1}$ and $h_{2l\pm1}$ already found.

We now proceed by induction. Suppose that we have proved that the formula (25.9) gives the solution (h) for $0 \leqq n < m$, where we are entitled by what proceeds to assume that $m \geqq 2l+2$. Since (h) satisfies (4.11), we obtain on taking n equal to l the relation

$$(25.11)\qquad h_{m+l}h_{m-l} = -h_{l+1}h_{l-1}h^2_m.$$

Now if $m \not\equiv \pm 1 \pmod{l}$, then $h_{m-1} = 0$ by the hypothesis of the induction. Hence $h_m = 0$ unless $m \equiv \pm 1 \pmod{l}$. Hence by the definition of λ_n, (25.9) is true if $n = m$ and $m \not\equiv \pm 1 \pmod{l}$.

Now assume that $m \equiv \pm 1 \pmod{l}$. Then on replacing m by $m-l$ in (25.11) we obtain the formula

$$(25.12)\qquad h_m = -h_{l-1}h_{l+1}h^2_{m-l}/h_{m-2l}.$$

For since $m \geqq 2l+2$, we have $m-2l > 0$ and $h_{m-2l} \neq 0$ by the hypothesis of the induction. We may now evaluate h_m by substituting in (25.12) for h_{m-l} and h_{m-2l} from (25.9). But we obtain in this manner (25.9) with n

replaced by m. Thus we have shown that if (25.9) holds for $0 \leqq n < m$, then it holds for $0 \leqq n < m+1$. Hence it is generally true by induction.

That conversely (25.9) is a solution of (1.1) is a trivial consequence of Theorem 24.1.

If we exclude from consideration the trivial solutions of (1.1) already discussed in which all except a finite number of terms are zero, we may summarize the results of Chapters IV, VI and the present sections as follows.

THEOREM 25.2. *Any non-trivial solution of*

$$(1.1) \qquad \omega_{m+n}\omega_{m-n} = \omega_{m+1}\omega_{m-1}\omega_n^2 - \omega_{n+1}\omega_{n-1}\omega_m^2$$

is equivalent to one of the following four solutions:

$$h_n = n; \quad h_n = \sin n\theta/\sin\theta; \quad h_n = \sigma(nu)/\sigma(u)^{n^2}; \quad h_n = \lambda_n c^{1-n\lambda_n}.$$

26. We have already remarked that the only non-trivial solutions of (1.1) with fourth term zero which can be divisibility sequences are those for which $l = 3$ so that h_3 is zero, but h_2 and h_4 are not zero. The formulas of Theorem 25.1 then give the general term of the sequence (h).

The question arises whether or not such a solution can be parameterized by elliptic functions, so that with a proper choice of invariants, $h_n = \psi_n(u)$. But (Halphen, *Traité des fonctions elliptiques,* Part I (1886), p. 96) we have in the notation of Chapter IV,

$$h_2 = \psi_2(u) = -\wp'(u)\,; \; h_3 = \psi_3(u)\,;$$
$$h_4 = \psi_4(u) = \wp'(u)\,(\wp'^4(u) - \psi_3(u)\,\wp''(u)).$$

Consequently, if $h_3 = 0$, it is necessary that $h_4 = -h_2^5$ for such a parameterization to be possible. But if this condition is satisfied, h_n reduces to $(-h)^{(n^2-1)/3}(n/3)$, so that (h) is equivalent to the Legendre symbol solution $(n/3)$. Now the Legendre symbol solution is equivalent to $(n/3)(-1)^{1-(n/3)n}$; for $(-1)^{n^2-1} = (-1)^{1-(n/3)n} = (-1)^{n(n-(n/3))} = +1$ if n is not divisible by three. But $(n/3)(-1)^{1-(n/3)n}$ is the special λ_n solution for $l = 3$ and $c = -1$; and this is evidently expressible as the Lucas solution

$$U_n = (\sin 2n\pi/3)/(\sin 2\pi/3)$$

satisfying the recurrence $U_{n+2} = U_{n+1} - U_n$. We may thus state the following theorem.

THEOREM 26.1. *If (h) is an elliptic divisibility sequence with the initial values $0, 1, h_2, 0, h_4$ where $h_2 h_4 \neq 0$, then (h) cannot be parameterized in terms of elliptic functions unless $h_4 = -h_2^5$. If this condition is satisfied, (h) is equivalent to the Lucas solution* $\sin(2n\pi/3)/\sin(2\pi/3)$.

VIII. Periodic Sequences.

27. We shall determine in this chapter all periodic elliptic sequences other than the special periodic sequences (λ) already discussed in Section **24** of the preceeding chapter. We shall be concerned here then with sequences (h) with $h_0 = 0$, $h_1 = 1$ and not both h_2 and h_3 zero. By Lemma 4.1 of Chapter IV, if two consecutive terms of such a sequence vanish, then all terms vanish beyond the third, and we have the trivial solution $0, 1, h_2, 0, 0, 0, \cdots$ of period one. It is easy to see conversely that this solution is the only one of period one. We shall now show that every other periodic sequence is purely periodic.

THEOREM 27.1. *Let (h): $0, 1, h_2, h_3, \cdots$ be a solution of (1.1) in which no two consecutive terms vanish. Then if (h) is periodic, (h) is purely periodic.*

Proof. Since if h_2 is zero, h_3 is not zero, and the conditions for periodicity in this case are trivial, it suffices to show that if no two consecutive terms of (h) vanish, then the assumptions

(27.1) $$h_{n+\kappa} = h_n, \qquad n \geqq a \geqq 1, \qquad \kappa \geqq 2;$$

(27.2) $$h_{a-1+\kappa} \neq h_{a-1};$$

(27.3) $$h_2 \neq 0;$$

lead to a contradiction. (These conditions simply state that (h) becomes periodic with period $\kappa > 1$ after a non-periodic terms.)

We shall begin by showing that

(27.4) $$h_\kappa = 0.$$

For, taking $m = a + \kappa - 1$ and $n = a + 1$ in the basic recursion (4.11), we obtain from (27.1), $h_{2a+2}h_\kappa = 0$. Hence either h_κ, or $h_{2a+2} = 0$. But if $h_{2a+2} = 0$, then on taking $m = 2a + 2 + \kappa$ and $n = \kappa$ in (4.11), we obtain $0 = h_{2a+1}h_{2a+3}h_\kappa^2$. Since neither h_{2a+1} nor h_{2a+3} can vanish, $h_\kappa = 0$.

We next show that

(27.5) $$\text{Either } h_a = 0 \text{ or } h_{a+1} = 0.$$

For taking $m = a + \kappa$ and $n = a$ in (4.11) we find that

$$h_{2a}h_\kappa = h_{a+1}h_a^2(h_{a-1+\kappa} - h_{a-1}).$$

Hence (27.5) follows from (27.4) and (27.2). Since $h_2 \neq 0$, it follows from (27.5) and (4.6) that either $h_{2a} = 0$ or $h_{2a+2} = 0$. Hence

(27.6) $$h_{2a+1} \neq 0.$$

We can now show that

$$(27.7) \qquad h_{a+1} = 1, \qquad h_{\kappa-1} = -1.$$

For taking $m = a + \kappa + 1$ and $n = a$ in (4.11) and reducing by (27.1) and (4.5), we find that $h_{2a+1}h_{\kappa+1} = h_{2a+1}$. Hence by (27.6), $h_{\kappa+1} = 1$. Next, taking $m = a + 1 + 2\kappa$ and $n = a$ in (4.11), we obtain the formula $h_{2a+1}h_{2\kappa+1} = h_{2a+1}$. Hence $h_{2\kappa+1} = 1$. But by (4.5), $h_{2\kappa+1} = -h_{\kappa-1}h^2{}_{\kappa+1}$, completing the proof of (27.7).

Next,

$$(27.8) \qquad h_{a-1+\kappa} = 0.$$

For taking $m = a - 1 + \kappa$ and $n = \kappa$ in (4.11), we obtain by (27.4) and (27.7), $h_{a-1+2\kappa}h_{a-1} = h^2{}_{a-1+\kappa}$. Since by (27.1) and (27.2), $h_{a-1+2\kappa} = h_{a-1+\kappa} \neq h_{a-1}$, (27.8) follows.

Finally,

$$(27.9) \qquad h_{a+1} = 0; \quad h_a \neq 0; \quad h_{a+2} \neq 0; \quad h_{a-1} \neq 0.$$

For by (27.5), either h_{a+1} or h_a equals zero. But $h_a = 0$ implies $h_{a+\kappa} = 0$ contrary to (27.8). Hence $h_{a+1} = 0$. Consequently $h_a \neq 0$ and $h_{a+2} \neq 0$; $h_{a-1} \neq 0$ by (27.2) and (27.8).

We may obtain a contradiction of (27.9) as follows. Take $m = a + 1 + \kappa$ and $n = a - 1 + \kappa$ in (4.11). Then $h_m = h_n = 0$ so that $h_{m+n}h_{m-n} = h_{2a}h_2 = 0$. Hence by (27.3), $h_{2a} = 0$. But by (4.6),

$$0 = h_{2a}h_2 = h_a(h_{a+2}h^2{}_{a-1} - h_ah^2{}_{a+1}) \text{ or } h_ah_{a+2}h^2{}_{a-1} = 0,$$

contradicting (27.9) and completing the proof of the theorem.

28. We have already shown the existence of periodic solutions of (1.1) with h_2 or h_3 zero of periods one, three, four, six and eight. The three theorems which follows are useful for deciding whether or not a given sequence is a periodic solution of (1.1). They may be proved either by mathematical induction of more briefly, by using the elliptic function representation theorem of Chapter IV.

THEOREM 28.1. *Let $(h): 0, 1, h_2, h_3, \cdots$ be a general solution of (1.1), so that neither h_2 nor h_3 is zero. Then any one of the following three sets of conditions is necessary and sufficient for (h) to be periodic with period κ:*

$$(28.1) \qquad \begin{array}{lll} \text{(i)} & h_{n+\kappa} = h_n & (n = 0, 1, \cdots, \kappa) \\ \text{(ii)} & h_{\kappa-n} = -h_n & (n = 0, 1, \cdots, \kappa) \\ \text{(iii)} & h_{\kappa/2+n} = -h_n & (\kappa \text{ even}; n = 0, 1, \cdots, \kappa/2). \end{array}$$

THEOREM 28.2. *Let $h_0, h_1, \cdots, h_\kappa$ be a set of $\kappa + 1$ numbers satisfying the conditions* (28.1) (ii) *or* (28.1) (iii), *and also satisfying the basic recursion* (4.11) *for $m + n \leqq \kappa$. Then if κ_n denotes the least positive residue of n modulo κ, and if h_n is defined to be h_{κ_n} for $n \geqq 0$, then* (h) *is a periodic solution of* (1.1) *with period κ.*

THEOREM 28.3. *If* (h) *is any integral general elliptic sequence and if m is an integral modulus prime to both h_2 and h_3, then the previous two theorems hold if the periodicity is understood to mean numerical periodicity modulo m and if the equalities in the conditions* (26.1) *are replaced by congruences modulo m.*

To illustrate the theorems, suppose that we start with the initial values $h_0 = 0$, $h_1 = 1$, $h_2 = b \neq 0$, $h_3 = 1$ and $h_4 = 0$ and compute from (4.5) and (4.6) $h_5 = -1$, $h_6 = -b$, $h_7 = -1$ and $h_8 = 0$. Then the nine numbers $0, 1, b, 1, 0, -1, -b, -1, 0$, satisfy (28.1) (ii) for $\kappa = 8$. They therefore define a periodic solution of (1.1) of period eight which is an elliptic divisibility sequence if b is an integer. It is easy to prove that any elliptic divisibility sequence with $h_2 h_3 \neq 0$ and $h_4 = 0$ is equivalent to this periodic solution.

Again, let us start with the initial values $h_0 = 0$, $h_1 = 1$, $h_2 = 1$, $h_3 = -1$, $h_4 = -1$. We find that $h_5 = 0$. Hence (28.1) (ii) is satisfied with $\kappa = 5$. If we start with the initial values $0, 1, b, b, 1$ we find that $h_5 = 0$, $h_6 = -1$, $h_7 = -b$, $h_8 = -b$, $h_9 = -1$, $h_{10} = 0$. Hence (28.1) (ii) is satisfied with $\kappa = 10$, and we have two periodic solutions of (1.1) of periods five and ten, respectively.

We shall show in the next section that there are essentially no other periodic elliptic sequences.

29. A sequence (h) will be called a "normal solution" of (1.1) if

(29.1) $$h_{m+n} h_{m-n} = h_{m+1} h_{m-1} h_n^2 - h_{n+1} h_{n-1} h_m^2, \qquad m \geqq n \geqq 1;$$

(29.2) $$h_0 = 0,\ h_1 = 1;\ h_2, h_3, h_4 \text{ and } h_4/h_2 \text{ integers};$$

(29.3) $$(h_3\, h_4) = 1.$$

By Theorem 6.1 of Chapter III, if (h) is normal

(29.4) $$(h_n, h_{n+1}) = 1, \qquad (n = 1, 2, 3, \cdots)$$

and by Theorem 6.4,

(29.5) $$(h_n, h_m) = h_{(n,m)}.$$

Every purely periodic elliptic divisibility sequence is normal, for if (h)

is purely periodic with period $\kappa \geqq 2$, then $h_{2\kappa+1} = h_1 = 1$. Consequently $(h_3, h_4) = 1$ by Theorem 6.1.

Let (h) be any normal solution. Then if

(29.5) $$h_\rho = 0 \quad \text{but} \quad h_n \neq 0, \qquad 0 < n < \rho,$$

then (h) is said to be of rank ρ.

THEOREM 29.1. *If (h) is a normal solution of (1.1) of rank ρ, then (h) is purely periodic and its period is either ρ or 2ρ.*

Proof. Let $0 \leqq n \leqq \rho$, and take $m = n + \rho$ in (29.1). Then $h_{m+1}h_{m-1}h_n^2 = h_{n+1}h_{n-1}h_m^2$. But by (29.4), h_n is prime to $h_{n+1}h_{n-1}$. Hence h_n^2 divides h_m^2. Similarly, h_m^2 divides h_n^2. Hence $h_{n+\rho} = \pm h_n$, $(0 \leqq n \leqq \rho)$, and it is easily shown that either the plus sign or the minus sign must be taken with every n according as $h_{\rho+1} = +1$ or $h_{\rho+1} = -1$. Hence by Theorem 28.1, (h) is purely periodic with period ρ or 2ρ.

THEOREM 29.2. *If (h) is purely periodic, its rank is less than six.*

In other words, integral periodic elliptic sequences can have only the periods 1, 2, 3, 4, 5, 6, 8 or 10. That each of these periods may actually occur has already been demonstrated. The proof of Theorem 29.2 rests on a series of lemmas which we establish in the next section. The proof of the theorem concludes the section and chapter.

30. LEMMA 30.1. *If (h) is any solution of (1.1) and $m \geqq n \geqq p > 0$, then*

(30.1) $$h_{m+n}h_{m-n}h_p^2 = h_{m+p}h_{m-p}h_n^2 - h_{n+p}h_{n-p}h_m^2.$$

This result easily follows on substituting for $h_{m+p}h_{m-p}$ and $h_{n+p}h_{n-p}$ on the right of (30.1) their expressions obtained from (29.1).

LEMMA 30.2. *Let (h) be an elliptic divisibility sequence and h_r any non-vanishing term of (h). Then if $k_n = h_{nr}/h_r$, $(n = 0, 1, 2, \cdots)$ (k) is an elliptic divisibility sequence. Furthermore if (h) is normal, so is (k).*

For taking $p = r$, $m = mr$ and $n = nr$ in (30.1) and dividing by h_4^4, we find that (k) satisfies (1.1). (k) is evidently integral and $k_0 = 0$, $k_1 = 1$ while k_4/k_2 is an integer. Hence (k) is an elliptic divisibility sequence. Now if (h) is normal, $(h_{3r}, h_{4r}) = h_r$ by (29.5). Hence $(k_3, k_4) = 1$ and (k) is normal.

LEMMA 30. 3. *If p is a prime greater than five and (h) is normal, then h_p is never zero.*

Proof. Since p is a prime, if $h_p = 0$, (h) is of rank p and hence (h) is purely periodic with period p or $2p$ by Theorem 29. 1. Since $(h_3, h_4) = 1$, (h) is purely periodic modulo three and hence p must be divisible by the rank of apparition of three in (h). But it was shown in Chapter III that $\rho \leqq 7$. Hence p equals seven. But if $h_7 = 0$, then $h_6 = \pm 1$ and $h_8 = \pm 1$ by (29. 5). Hence $h_2 = \pm 1$, $h_4 = \pm 1$ and $h_3 = \pm 1$. Since $h_5 = h_4 h_2^3 - h_3^3 \neq 0$, $h_5 = \pm 2$. But then $h_7 = h_5 h_3^3 - h_2 h_4^3 = \pm 2 \pm 1 \neq 0$. This contradiction completes the proof of the lemma.

LEMMA 30. 4. *If ρ is the rank of (h), then ρ can contain no prime factor other than two, three or five.*

For if p is any prime factor of ρ, write $\rho = pq$. Then $h_q \neq 0$ by the definition of rank. Hence if $k_n = h_{nq}/h_q$, (k) is a normal sequence of rank p by Lemma 30. 2. Hence by Lemma 30. 3, p equals two, three or five.

LEMMA 30. 5. *Let (h) be a normal sequence of rank ρ. Then ρ is not equal to any one of the following numbers:*

$$6, 8, 9, 10, 15, 25. \tag{30. 2}$$

The proof proceeds by examination of cases; it suffices to give two examples. Suppose that $\rho = 6$. Then $h_5 = \pm 1$, $h_n \neq 0$, $0 < n < 6$ and $h_6 h_2 = h_3(h_5 h_2^2 - h_4^2) = 0$. Hence $h_5 = \pm 1$, $h_4 = \pm h_2$. But $h_5 = h_4 h_2^3 - h_3^3$. Hence one or the other of the diophantine equations $X^4 = 1 + Y^3$, $X^4 + 1 = Y^3$ must have non-zero integral solutions. But it is easily seen that neither has non-zero integral solutions. Hence $\rho \neq 6$.

Now suppose that $\rho = 10$. Then $h_q = \pm 1$, so $h_8 = \pm 1$, and since $h_{50} = 0$, $h_{49} = \pm 1$ so $h_7 = \pm 1$. Now $0 = h_{10} h_2 = h_5(h_7 h_4^2 - h_3 h_6^2)$. Hence $h_4^2 = h_6^2$, $h_3 = \pm 1$, $h_6 = \pm h_4$. Next, $h_6 h_2 = h_3(h_5 h_2^2 - h_4^2)$. Hence $h_4 \mid h_2$, $h_4 = \pm h_2$. But then $h_6 h_2 = \pm h_2^2 = h_3 h_2^2(h_5 - 1)$. Hence $h_5 - 1 = \pm 1$, and since $h_5 \neq 0$, $h_5 = 2$. But $h_9 = h_6 h_4^3 - h_3 h_5^3$. Hence $\pm 1 = \pm h_4^4 = 8$ or $h_4^4 = 7$ or 9 which is impossible. Hence $\rho \neq 10$. The other cases may be disposed of similarly.

LEMMA 30. 6. *Let (h) be a normal sequence of rank ρ. Then ρ is not divisible by any one of the numbers* (30. 2).

For let m be any one of the numbers (30. 2) and assume that $\rho = lm$,

$l \geqq 1$. Then $h_l \neq 0$ and $k_n = h_{ln}/h_l$ defines a normal sequence (k) of rank m contrary to Lemma 30. 5.

Proof of Theorem 29. 2. Let (h) be normal of rank ρ. By Lemma 30. 4, the only prime factors of ρ are two, three and five and by Lemma 30. 6, ρ is not divisible by 2^2, 3^2, 5^2 or 2×3, 2×5, 3×5. Hence ρ must equal two, three, four or five.

IX. Conclusion: Lucas' Conjecture.

31. The results obtained in Chapters VI and VII make it clear that the only solutions of (1. 1) that can be related to solutions of linear recurrences of order three or four are the general elliptic function solutions. Now the arithmetical behavior of a sequence of integers $(W): W_0, W_1, W_2, \cdots$ defined recursively by

$$W_{n+3} = PW_{n+2} + QW_{n+1} + RW_n$$

or

$$W_{n+4} = PW_{n+3} + QW_{n+2} + RW_{n+1} + TW_n$$

P, Q, R, T fixed integers, is well known. (Carmichael [1], Ward [1].) First of all, such a sequence is only exceptionally a divisibility sequence (Hall [1], Ward [2]), and if it is a divisibility sequence, the rank of any prime p in it divides $p(p^3 - 1)$ or $p(p^4 - 1)$ according as the recursion is of order three or four. Since there exist elliptic sequences in which the rank of every prime is five and since there are an infinite number of primes p such that $p_3 = 1$ is not divisible by five, no direct connection with recurrences of order three seems possible. In particular, there cannot exist a formula $h_n = K^{a_n} W_n$ analogous to Lucas' $h_n = q^{(1-n)/2} U_n$.

If (h) is singular and hence essentially a Lucas function, the rank of apparition of any prime in (h) may be shown to divide $p(p^4 - 1)$. But this is not true if (h) is non-singular. For consider the sequence with the initial values $0, 1, 1, 1, 5$. We find that $h_5 = 4$, $h_6 = -21$, and $h_7 = -121$. Hence the rank of apparition of the prime 11 is 7. But $11 \cdot (11^4 - 1) = 2^4 \cdot 3 \cdot 5 \cdot 11 \cdot 61$ is not divisible by 7.

If (W) is not a divisibility sequence, the prospects are even worse, for two consecutive terms of such a sequence may be divisible by a prime p without having almost all terms of (W) divisible by p, contrary to Theorem 6. 1.

Although the analogy between an elliptic sequence (h) and a Lucas sequence (U) is a close one, I should like to point out in concluding one

very significant difference. For a Lucas sequence, (and more generally for any linear divisibility sequence) it is possible to name in advance terms which will certainly be divisible by a given prime p; for example $U_{p\pm1}$ for the Lucas function proper. Consequently, the rank of apparition of p is arithmetically restricted since it must divide either $p-1$ or $p+1$. But for the general elliptic sequence (h), computational experiments disclose no such simple arithmetical connections between a prime and its rank of apparition; it appears to be impossible to name in advance a particular h_k which will be divisible by a given prime p.

CALIFORNIA INSTITUTE OF TECHNOLOGY.

BIBLIOGRAPHY

E. T. Bell.

[1] *Bulletin of the American Mathematical Society*, vol. 29 (1923), pp. 401-406.

R. D. Carmichäel.

[1] *Quarterly Journal of Mathematics* vol. 48 (1920), pp. 343-372.

M. Hall.

[1] *American Journal of Mathematics*, vol. 38 (1936), pp. 577-584.

E. Lucas.

[1] and [2] *American Journal of Mathematics*, vol. 1 (1878), pp. 184-240; 289-321.

M. Ward.

[1] *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 153-165.

[2] *Ibid.*, vol. 44 (1938), pp. 68-86.

[3] *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 843-845.

METHODS OF SUMMABILITY WHICH EVALUATE SEQUENCES OF ZEROS AND ONES SUMMABLE C_1.*

By Ralph Palmer Agnew.

1. Introduction. It is the object of this paper to show if A is a method of summability belonging to a general class, and if each divergent sequence of zeros and ones which is evaluable by the Cesàro arithmetic mean method C_1 is evaluable A to the same value, then A must be regular.

Let $a_1(t), a_2(t), \cdots$ be a sequence of complex-valued functions defined over a set T, in a metric or Hausdorff space, having a limit point t_0. These functions determine a sequence-to-function transformation

$$\sigma(t) = \sum_{k=1}^{\infty} a_k(t) s_k \tag{A}$$

by means of which a sequence $s_1, s_2, \cdots$ of complex numbers is evaluated to σ if the series in (A) converges, when t is in T, to a function $\sigma(t)$ such that $\sigma(t) \to \sigma$ as $t \to t_0$. We shall use the symbol $As_n = \sigma$ to abbreviate the statement that A evaluates the sequence s_n to the value σ. The transformation A is regular if $As_n = \lim s_n$ whenever $\lim s_n$ exists. The standard necessary and sufficient conditions for regularity of A are

$$\sum_{k=1}^{\infty} | a_k(t) | < \infty \qquad t \,\varepsilon\, T \tag{1}$$

$$\limsup_{t \to t_0} \sum_{k=1}^{\infty} | a_k(t) | = M < \infty \tag{2}$$

$$\lim_{t \to t_0} a_k(t) = 0 \qquad k = 1, 2, 3, \cdots \tag{3}$$

$$\lim_{t \to t_0} \sum_{k=1}^{\infty} a_k(t) = 1. \tag{4}$$

A familiar subfamily of transformations of the form A is that for which T is the set of positive integers and $t_0 = +\infty$. Such transformations have the form

$$\sigma_n = \sum_{k=1}^{\infty} a_{nk} s_k, \tag{A$'$}$$

* Received September 5, 1946; Presented to the American Mathematical Society, August 19, 1946.

a sequence s_n being evaluable A' to σ if the series converge to numbers $\sigma_1, \sigma_2, \cdots$ such that $\lim \sigma_n = \sigma$.

A sequence s_n is evaluable to σ by C_1 (the Cesàro method of order 1) if

$$\lim_{n\to\infty} (s_1 + s_2 + \cdots + s_n)/n = \sigma. \tag{5}$$

THEOREM 1. *If $As_n = C_1 s_n$ whenever s_n is a divergent sequence of 0's and 1's for which $C_1 s_n$ exists, then A must be regular.*

Actually, we shall prove the following stronger theorem with weaker hypotheses.

THEOREM 2. *If $As_n = \frac{1}{2}$ whenever s_n is a sequence of zeros and ones ending with $0, 1, 0, 1, 0, 1, \cdots$, and if $As_n = 0$ whenever s_n is a divergent sequence of zeros and ones for which $C_1 s_n = 0$, then A must be regular.*

That Theorem 2 is in fact stronger than Theorem 1 is a consequence of the following trivial example. The transformation

$$\sigma_n = (s_3 + s_6 + s_9 + \cdots + s_{3n})/n \tag{6}$$

satisfies the hypothesis of Theorem 2, but not that of Theorem 1.

2. Remarks. Borel [4] discovered the fact that almost all sequences of zeros and ones are evaluable C_1 to $1/2$. Hence the hypothesis of Theorem 1 implies that almost all sequences of zeros and ones are evaluable A to $1/2$. But the hypothesis that almost all sequences of zeros and ones are evaluable A to $1/2$ does not imply that A is regular. In fact, the transformation

$$\sigma_n = (1/n) \sum_{k=1}^{n} s_k + \sum_{k=n+1}^{\infty} (-1)^k (1/k) s_k \tag{7}$$

with matrix

$$\begin{matrix} 1 & \frac{1}{2} & -\frac{1}{3} & \frac{1}{4} & -\frac{1}{5} & \frac{1}{6} & \cdots \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{3} & \frac{1}{4} & -\frac{1}{5} & \frac{1}{6} & \cdots \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{4} & -\frac{1}{5} & \frac{1}{6} & \cdots \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & -\frac{1}{5} & \frac{1}{6} & \cdots \\ \cdots & & & & & & \end{matrix} \tag{8}$$

is nonregular since, for each n, the series whose terms are the absolute values of the elements in the n-th row is divergent. The Borel theorem cited above says that, for almost all sequences of zeros and ones, the first term in the

right member of (7) converges to $\frac{1}{2}$ as $n \to \infty$. Denote the second term in the right member of (7) by σ'_n and write it in the form

$$\sigma'_n = \tfrac{1}{2} \sum_{k=n+1}^{\infty} (-1)^k (1/k) r_k(t) + \tfrac{1}{2} \sum_{k=n+1}^{\infty} (-1)^k (1/k) \tag{9}$$

where t is the number in the interval $0 \leqq t \leqq 1$ equal to the dyadic number $.s_1 s_2 s_3 \cdots$ and $r_n(t) = 2s_n - 1$ is the n-th Rademacher function. We now use Rademacher's elegant theorem (See Rademacher [6] or Kacmarz and Steinhaus [5]) that *if* $\Sigma \mid a_k \mid^2 < \infty$, *then* $\Sigma\, a_k r_k(t)$ *converges for almost all* t. This implies that, for almost all t, the first term in the right member of (9) converges to 0 as $n \to \infty$. The second term converges to 0. Hence, for almost all t, $\sigma'_n \to 0$, $\sigma_n \to \frac{1}{2}$, and the transformation (7) has the property in question.

3. A lemma on sums of complex numbers. Before passing to the proof of Theorem 2, we state and prove a simple lemma which we shall use.

LEMMA 1. *If* $z_1, z_2, \cdots, z_n$ *is a finite set of complex numbers, then there is a set* $\theta_1, \theta_2, \cdots, \theta_n$ *of numbers, of which each element is* 0 *or* 1, *such that*

$$\sum_{k=1}^{n} \mid z_k \mid \leqq 6 \mid \sum_{k=1}^{n} \theta_k z_k \mid . \tag{10}$$

If Q is a suitably selected one of the four quadrants into which the real and imaginary axes cut the complex plane, and if $\theta_k = 1$ when z_k lies in the closure of Q and $\theta_k = 0$ otherwise, then

$$\sum_{k=1}^{n} \mid z_k \mid \leqq 4 \sum_{k=1}^{n} \mid \theta_k z_k \mid . \tag{11}$$

Since the members of (10) remain unchanged when the z's are all multiplied by the same power of i, we may suppose Q is the first quadrant. Then, when $z_k = a_k + ib_k$,

$$\begin{aligned} \sum_{k=1}^{n} \mid \theta_k z_k \mid &\leqq \sum_{k=1}^{n} (\theta_k a_k + \theta_k b_k) \\ &\leqq 2^{1/2} \mid \sum_{k=1}^{n} (\theta_k a_k + i\theta_k b_k) \mid = 2^{1/2} \mid \sum_{k=1}^{n} \theta_k z_k \mid . \end{aligned} \tag{12}$$

Since $4(2)^{1/2} = 32^{1/2} < 6$, the conclusion of Lemma 1 follows.

4. Proof of the theorems. Let A satisfy the hypothesis of Theorem 2. To prove regularity of A, we show that (1), (2), (3), and (4) must hold.

To prove (4) let s'_k and s''_k be respectively the sequences $0, 1, 0, 1, \cdots$ and $1, 0, 1, 0, \cdots$. Then $C_1 s'_k = \frac{1}{2}$ and $C_1 s''_k = \frac{1}{2}$; hence

$$(13) \qquad \lim_{t\to t_0} \sum_{j=1}^{\infty} a_{2j-1}(t) = \tfrac{1}{2}, \qquad \lim_{t\to t_0} \sum_{j=1}^{\infty} a_{2j}(t) = \tfrac{1}{2}$$

and this implies (4).

To prove (3), let q be a positive integer. Let s'_k and s''_k be the sequence $0, 1, 0, 1, 0, 1, \cdots$ except that $s'_q = 1$ and $s''_q = 0$. Then $C_1 s'_k = C_1 s''_k = \frac{1}{2}$. Hence

$$(14) \qquad a_q(t) = \sum_{k=1}^{\infty} a_k(t) s'_k - \sum_{k=1}^{\infty} a_k(t) s''_k \to \tfrac{1}{2} - \tfrac{1}{2} = 0$$

as $t \to t_0$, and (3) is proved.

To prove (1) suppose, intending to establish a contradiction of our hypothesis, that (1) fails. Then, for some fixed t in T,

$$(15) \qquad \sum_{k=1}^{\infty} | a_k(t) | = \infty.$$

Then, as the author [1] has proved, there is a (necessarily divergent) sequence $s_1, s_2, \cdots$ of zeros and ones such that $C_1 s_k = 0$ and $\Sigma\, a_k(t) s_k$ is not convergent. This contradicts our hypothesis and proves that (1) must hold. The referee pointed out to the author that existence of the sequence $s_1, s_2, \cdots$ can also be inferred from a theorem of Auerbach [3] on subseries of divergent series of positive terms.

To prove (2) suppose, intending to establish a contradiction, that (2) fails. Then

$$(16) \qquad \limsup_{t\to t_0} \sum_{k=1}^{\infty} | a_k(t) | = \infty.$$

Let $t_1, t_2, \cdots$ be a sequence of points of t such that $t_n \to t_0$ and

$$(17) \qquad \lim_{n\to\infty} \sum_{k=1}^{\infty} | a_k(t_n) | = \infty.$$

To simplify typography, let

$$(18) \qquad a(n, k) = a_k(t_n) \qquad\qquad n, k = 1, 2, 3, \cdots$$

so that (17) becomes

$$(19) \qquad \lim_{n\to\infty} \sum_{k=1}^{\infty} | a(n, k) | = \infty.$$

We have also the conditions

$$(20) \qquad \lim_{n\to\infty} a(n,k) = 0, \qquad \sum_{k=1}^{\infty} | a(n,k) | < \infty$$

for each k and n, respectively. These conditions enable us to choose, in the order listed, integers

$$(21) \qquad \alpha_1, n_1, \beta_1, \alpha_2, n_2, \beta_2, \alpha_3, n_3, \beta_3, \cdots$$

satisfying the conditions below. Let $\alpha_1 = 1$. Choose n_1 and then β_1 such that

$$(22) \qquad \sum_{k=\alpha_1}^{\infty} | a(n_1,k) | > 6, \qquad \sum_{k=\alpha_1}^{\beta_1-1} | a(n_1,k) | > 6$$

and then choose α_2 such that

$$(23) \qquad \alpha_2 > 1 + \beta_1, \qquad \sum_{k=\alpha_2}^{\infty} |a(n_1,k)| < 1.$$

When n_{p-1}, β_{p-1}, and α_p have been defined, choose n_p such that $n_p > n_{p-1}$ and

$$(24) \qquad \sum_{k=1}^{\alpha_p-1} | a(n_p,k) | < p^{-1}. \qquad \sum_{k=\alpha_p}^{\infty} | a(n_p,k) | > 6p^2.$$

Then choose β_p such that

$$(25) \qquad \beta_p = \alpha_p + p\gamma_p, \qquad \sum_{k=\alpha_p}^{\beta_p-1} | a(n_p,k) | > 6p^2,$$

where γ_p is a positive integer; and then choose α_{p+1} such that

$$(26) \qquad \alpha_{p+1} > p + \beta_p, \qquad \sum_{k=\alpha_{p+1}}^{\infty} |a(n_p,k)| < p^{-1}.$$

For each $p = 1, 2, 3, \cdots$

$$(27) \qquad \sum_{k=0}^{p-1} \sum_{j=0}^{\gamma_p-1} | a(n_p, \alpha_p + k + jp) | = \sum_{k=\alpha_p}^{\beta_p-1} | a(n_p,k) | > 6p^2;$$

hence there must be an index k_p such that $0 \leqq k_p < p$ and

$$(28) \qquad \sum_{j=0}^{\gamma_p-1} | a(n_p, \alpha_p + k_p + jp) | > 6p.$$

Let $s_n = 0$ when $\alpha_p \leqq n < \alpha_{p+1}$ except that $s_n = 1$ for those values of n in the set

$$n = \alpha_p + k_p + jp \qquad\qquad j = 0, 1, \cdots, \gamma_p - 1$$

so selected that

$$(30) \qquad \left|\sum_{j=0}^{\gamma_p-1} a(n_p, \alpha_p + k_p + jp) s(\alpha_p + k_p + jp)\right| > p,$$

this selection of the s's being possible because of (28) and Lemma 1. Then

$$(31) \qquad \left|\sum_{k=\alpha_p}^{\alpha_{p+1}-1} a(n_p, k) s_k\right| > p.$$

If $k_1, k_2, \cdots$ denote in increasing order the values of k for which $s_k = 1$, the construction of the sequence s_k implies that $k_{q+1} - k_q \to \infty$ as $q \to \infty$; this, with the fact that s_k is always zero or one, implies (5) with $\sigma = 0$; hence $C_1 s_n = 0$. But, for each $p = 2, 3, \cdots$,

$$(32) \qquad \left|\sum_{k=1}^{\infty} a_k(t_{np}) s_k\right| = \left|\sum_{k=1}^{\infty} a(n, k) s_k\right|$$

$$\geqq -\sum_{k=1}^{\alpha_p-1} |a(n_p, k)| s_k + \left|\sum_{k=\alpha_p}^{\alpha_{p+1}-1} a(n_p, k) s_k\right| - \sum_{k=\alpha_p}^{\infty} |a(n_p, k)| s_k$$

$$\geqq -p^{-1} + p - p^{-1}.$$

Thus $\Sigma a_k(t) s_k$ fails to have a limit as $t \to t_0$, As_k fails to exist, and (2) must hold. This completes the proof of Theorem 2 and hence also that of Theorem 1.

5. A variant of Theorem 2.

THEOREM 3. *If $As_n = \frac{1}{2}$ whenever s_n is a sequence of zeros and ones for which $C_1 s_n = \frac{1}{2}$, then A must be regular.*

We prove this theorem by showing that its hypotheses imply those of Theorem 2. Since each sequence ending with $0, 1, 0, 1, 0, 1, \cdots$ is evaluable C_1 to $\frac{1}{2}$, the first part of the hypothesis of Theorem 1 is implied. Now let s'_n be any sequence of zeros and ones evaluable C_1 to 0. Let s''_n be the sequence $0, 1, 0, 1, 0, 1, \cdots$ except that $s''_n = 0$ for each n for which $s'_n = 1$. Then $C_1 s''_n = \frac{1}{2}$ and $C_1\{s''_n + s'_n\} - \frac{1}{2}$. Therefore, since both s''_n and $s''_n + s'_n$ are sequences of zeros and ones, $As''_n = \frac{1}{2}$ and $A\{s''_n + s'_n\} = \frac{1}{2}$. Since A is linear, $A\{s'_n\} = A\{s''_n + s'_n - s''_n\} = A\{s''_n + s'_n\} - As''_n = \frac{1}{2} - \frac{1}{2} = 0$; thus the second part of the hypothesis of Theorem 2 is implied.

6. Conclusion. While the hypothesis of Theorem 1 implies that A is regular, it does not imply that A includes C_1, that is, that $As_n = C_1 s_n$ for

every sequence s_n for which $C_1 s_n$ exists. In fact, if $0 < r < 1$, then the Cesàro transformation C_r has the form A. But (see Andersen [2]) $C_r s_n = C_1 s_n$ whenever s_n is a bounded sequence for which $C_1 s_n$ exists, and C_r fails to include C_1.

CORNELL UNIVERSITY,
ITHACA, NEW YORK.

REFERENCES.

1. R. P. Agnew, "Subseries of series which are not absolutely convergent," *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 118-120.

2. A. F. Andersen, "Studier over Cesàro's summabilitetsmetode," *Thesis*, Copenhagen, 1921, 100 pp.

3. H. Auerbach, "Über die Vorzeichenverteilung in unendlichen Reihen," *Studia Mathematica*, vol. 2 (1930), pp. 228-230.

4. Borel, Émile, "Les probabilités dénombrables et leurs applications arithmétiques," *Rendiconti del Circolo Matematico di Palermo*, vol. 27 (1909) pp. 247-271.

5. S. Kaczmarz and H. Steinhaus, *Theorie der Orthogonalreihen*, Warszawa-Lwów, 1935, 298 pp., especially p. 126.

6. Rademacher, H., "Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen," *Mathematische Annalen*, vol. 87 (1922), pp. 112-138.

THE EXCEPTIONAL SIMPLE JORDAN ALGEBRAS.*

By R. D. Schafer.

A Jordan algebra is a non-associative algebra $\mathfrak{A}$ of finite order over a field $\mathfrak{F}$ which is commutative and in which

$$(1) \qquad a^2(ab) = a(a^2b)$$

holds for all a, b in $\mathfrak{A}$. The examples which come most easily to mind are the so-called Jordan algebras of linear transformations. Let $\mathfrak{M}_n$ be the algebra of all linear transformations on a vector space $\mathfrak{L}$ of dimension n over $\mathfrak{F}$, and $\mathfrak{A}$ be a linear subset of $\mathfrak{M}_n$ which is closed with respect to "quasi-multiplication"

$$(2) \qquad ab = \tfrac{1}{2}(a \cdot b + b \cdot a),$$

where $a \cdot b$ denotes the ordinary (associative) multiplication of transformations a, b in $\mathfrak{M}_n$. In this case $\mathfrak{A}$ with multiplication defined by (2) is a Jordan algebra over $\mathfrak{F}$ and is called a Jordan algebra of linear transformations ([3]).

A. A. Albert has recently developed a structure theory for Jordan algebras over nonmodular fields. In [4] he shows that the radical $\mathfrak{N}$ of $\mathfrak{A}$ is the unique maximal solvable ideal of $\mathfrak{A}$, and is even strongly nilpotent. The difference algebra $\mathfrak{A} - \mathfrak{N}$ is the direct sum of simple Jordan algebras, while the theory of simple Jordan algebras corresponds to the original work in 1934 of Jordan, von Neumann, and Wigner ([6]) and Albert ([1]): all simple Jordan algebras are Jordan algebras of linear transformations with the exception of a class of algebras which have degree 3 and order 27 over their centers. These algebras we call the exceptional simple Jordan algebras.

Simple Jordan algebras are central simple (i. e., simple for all scalar extensions) over their centers. It is shown in [4] that the exceptional central simple Jordan algebras $\mathfrak{A}$ over a nonmodular field $\mathfrak{F}$ have $\mathfrak{M}_3{}^8$ as split algebra. That is, there exists a scalar extension $\mathfrak{K}$ of finite degree over $\mathfrak{F}$ such that $\mathfrak{A}_{\mathfrak{K}}$ is the Jordan algebra $\mathfrak{M}_3{}^8$ of 3-rowed Hermitian matrices with elements in the unique Cayley-Dickson algebra $\mathfrak{D}$ over $\mathfrak{K}$ having norm form $\xi_1{}^2 + \xi_2{}^2 + \cdots + \xi_8{}^2$, multiplication in $\mathfrak{M}_3{}^8$ being given by (2) where $a \cdot b$ denotes the ordinary matrix product.

* Received April 20, 1947; Presented to the American Mathematical Society, April 26, 1947.

In this paper we show that these exceptional central simple algebras $\mathfrak{A}$ are reduced algebras (i. e., their primitive idempotents remain primitive under scalar extension), and that they consist of the J-symmetric elements of $\mathfrak{M}_3 \times \mathfrak{C}$, where $\mathfrak{C}$ is any Cayley-Dickson algebra over $\mathfrak{F}$, and J is the involution

$$(3) \qquad J: \quad a \to aJ = p\bar{a}'p^{-1}$$

of $\mathfrak{M}_3 \times \mathfrak{C}$, p a non-singular diagonal matrix in $\mathfrak{M}_3$. Multiplication in $\mathfrak{A}$ is that given by (2) where $a \cdot b$ denotes the (non-associative) multiplication in $\mathfrak{M}_3 \times \mathfrak{C}$. If $\mathfrak{F}$ is formally real, then there exist inequivalent Jordan algebras over $\mathfrak{F}$ with $\mathfrak{M}_3{}^8$ as split algebra.

1. The non-associative algebras $\mathfrak{M}_n \times \mathfrak{C}$. The Cayley-Dickson algebras $\mathfrak{C}$ over $\mathfrak{F}$ (of arbitrary characteristic) are central simple alternative algebras of degree 2 and order 8 over $\mathfrak{F}$ defined as follows: let $\mathfrak{Q}$ be a (generalized) quaternion algebra over $\mathfrak{F}$ with involution $q \to \bar{q} = t(q)1 - q$ for q in $\mathfrak{Q}$; then $\mathfrak{C} = \mathfrak{Q} + v\mathfrak{Q}$, $v^2 = \gamma \neq 0$ in $\mathfrak{F}$, $qv = v\bar{q}$, so that every element x of $\mathfrak{C}$ may be written in the form $x = q_1 + vq_2$. We have $x^2 - t(x)x + n(x)1 = 0$, and the correspondence

$$(4) \qquad x \to \bar{x} = t(x)1 - x = \bar{q}_1 - q_2$$

is an involution of $\mathfrak{C}$ satisfying

$$(5) \qquad x + \bar{x} = t(x)1, \qquad x\bar{x} = \bar{x}x = n(x)1.$$

Multiplication in $\mathfrak{C}$ has the explicit form

$$(6) \qquad (q_1 + vq_2)(q_3 + vq_4) = (q_1q_3 + \gamma q_4\bar{q}_2) + v(\bar{q}_1q_4 + q_3q_2)$$

for q_i in $\mathfrak{Q}$.

If the characteristic of $\mathfrak{F}$ is not two, there is a convenient normalization of the basis in a Cayley-Dickson algebra $\mathfrak{C}$. For then we may write $\mathfrak{C} = (1, u_1, \cdots, u_7)$ with $u_i{}^2 = \gamma_i \neq 0$ in $\mathfrak{F}$, $u_iu_j = -u_ju_i = \pm u_k$ for $i \neq j$. Now $\frac{1}{2}t(x\bar{y})$ is a symmetric bilinear form in the coefficients of elements x, y in $\mathfrak{C}$ such that the corresponding quadratic form is $n(x) = \frac{1}{2}t(x\bar{x})$. Written in matrix notation,

$$(7) \qquad \tfrac{1}{2}t(x\bar{y}) = xCy', \qquad n(x) = xCx',$$

where C is the non-singular diagonal matrix

$$(8) \qquad C = \operatorname{diag}\{1, -\gamma_1, \cdots, -\gamma_7\}.$$

We denote by $\mathfrak{D}$ the unique Cayley-Dickson algebra over $\mathfrak{F}$ with $\gamma_i = -1$ $(i = 1, \cdots, 7)$. For this algebra $C = I$ and $\frac{1}{2}t(x\bar{y}) = xy'$; we denote the norm form xx' by $d(x)$. Right and left multiplications R_x and L_x of elements of $\mathfrak{D}$ have the property

(9) $$R_x' = R_{\bar{x}}, \qquad L_x' = L_{\bar{x}}.$$

In the solution of our problem we are led to consider the direct products $\mathfrak{M}_n \times \mathfrak{C}$—that is, the (non-associative) algebras of all n-rowed square matrices with elements in $\mathfrak{C}$, a Cayley-Dickson algebra over $\mathfrak{F}$. In particular, we require the case $n = 3$.

The algebras $\mathfrak{M}_n \times \mathfrak{C}$ are involutorial central simple algebras of order $8n^2$ over $\mathfrak{F}$. For $(\mathfrak{M}_n \times \mathfrak{C})_{\mathfrak{K}} = \mathfrak{M}_n \times \mathfrak{C}_{\mathfrak{K}}$ is certainly simple for any scalar extension $\mathfrak{K}$ of $\mathfrak{F}$. Moreover, the correspondence $a \to \bar{a}'$ in $\mathfrak{M}_n \times \mathfrak{C}$ induced by $m \to m'$ in $\mathfrak{M}_n$, $x \to \bar{x}$ in $\mathfrak{C}$, is an involution. We are concerned with the more general involution (3) induced in $\mathfrak{M}_n \times \mathfrak{C}$ by $x \to \bar{x}$ in $\mathfrak{C}$ and

(10) $$m \to pm'p^{-1} \text{ in } \mathfrak{M}_n,$$

where p is any non-singular diagonal matrix in $\mathfrak{M}_n$. To see that (3) is an involution we write any elements a, b of $\mathfrak{M}_n \times \mathfrak{C}$ as $a = (x_{ij})$, $b = (y_{jk})$ where x_{ij}, y_{jk} are in $\mathfrak{C}$ $(i, j, k = 1, \cdots, n)$. Then $ab = (x_{ij})(y_{jk}) = (\Sigma_j x_{ij} y_{jk})$, $(ab)J = p(\overline{\Sigma_j x_{ij} y_{jk}})' p^{-1} = p(z_{ik}) p^{-1}$ where $z_{ki} = \Sigma_j \bar{y}_{jk} \bar{x}_{ij}$. On the other hand, $bJ \cdot aJ = p(\bar{y}_{ij})' p^{-1} p(\bar{x}_{jk})' p^{-1} = p(s_{ij})(t_{jk}) p^{-1} = p(\Sigma_j s_{ij} t_{jk}) p^{-1}$ where $s_{ij} = \bar{y}_{ji}$, $t_{jk} = \bar{x}_{kj}$ so that $\Sigma_j s_{ij} t_{jk} = \Sigma_j \bar{y}_{ji} \bar{x}_{kj} = z_{ik}$, $(ab)J = bJ \cdot aJ$. Also $aJ^2 = \overline{paJ'p^{-1}} = p(\overline{p\bar{a}'p^{-1}})' p^{-1} = p(p^{-1}ap)p^{-1} = a$ since $p = \bar{p} = p'$ while (3) has already been shown to be an anti-automorphism.

If n is odd, any involution U of $\mathfrak{M}_n$ is cogredient with an involution of the form (10). For by [2], Theorem 10.11, we have $mU = tm't^{-1}$ for some non-singular $t = \pm t'$ in $\mathfrak{M}_n$. Since n is odd, t would be singular in case it were skew. Hence t, being symmetric, is congruent to a non-singular diagonal matrix p in $\mathfrak{M}_n$, and U is cogredient with (10). That is, U and (10) are merely different representations of the same abstract involution of $\mathfrak{M}_n$.

If $n = 2k$ is even, an involution U of $\mathfrak{M}_n$ may be either cogredient with (10) or, in case $t = -t'$, cogredient with

(11) $$m \to gm'g^{-1} \text{ where } g = \operatorname{diag}\{f, \cdots, f\}, \qquad f = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

for in this case t is congruent to g. Since any element of $\mathfrak{M}_{2k} = \mathfrak{M}_k \times \mathfrak{M}_2$ may be written as a matrix (m_{ij}), m_{ij} in $\mathfrak{M}_2$, $(i, j = 1, \cdots, k)$, the involution

(11) is induced in $\mathfrak{M}_n$ by transposition in $\mathfrak{M}_k$ and the involution $m_{ij} \to f m_{ij} f^{-1}$ in $\mathfrak{M}_2$.

Corresponding to the quaternion subalgebra $\mathfrak{Q}$ of $\mathfrak{C} = \mathfrak{Q} + v\mathfrak{Q}$ there is an associative subalgebra $\mathfrak{W} = \mathfrak{M}_n \times \mathfrak{Q}$ of $\mathfrak{M}_n \times \mathfrak{C}$. Moreover, $\mathfrak{M}_n \times \mathfrak{C} = \mathfrak{W} + u\mathfrak{W}$, where $u = \operatorname{diag}\{v, \cdots, v\}$. Multiplication in $\mathfrak{M}_n \times \mathfrak{C}$ may be written

$$(12) \qquad (w_1 + uw_2)(w_3 + uw_4) = (w_1 w_3 + \gamma \overrightarrow{w_2 \bar{w}_4}) + u(\bar{w}_1 w_4 + \overrightarrow{\bar{w}_2 \bar{w}_3})$$

for w_i in $\mathfrak{W}$, $w \to \bar{w}$ being the correspondence induced in $\mathfrak{W}$ by $q \to \bar{q}$ in $\mathfrak{Q}$. This multiplication resembles (6), but differs in that $w \to \bar{w}$ is not an involution of $\mathfrak{W}$.

2. Exceptional central simple Jordan algebras. It is known ([4]) that the set of Hermitian elements (i. e., $a = \bar{a}'$) of $\mathfrak{M}_3 \times \mathfrak{D}$ with multiplication defined by (2) is a central simple Jordan algebra of degree three and order 27 over $\mathfrak{F}$; we customarily write $\mathfrak{M}_3{}^8$ for this algebra. Any exceptional central simple Jordan algebra $\mathfrak{A}$ over a nonmodular field $\mathfrak{F}$ has $\mathfrak{M}_3{}^8$ as split algebra: there exists a scalar extension $\mathfrak{K}$ of $\mathfrak{F}$ such that $\mathfrak{A}_{\mathfrak{K}} \cong (\mathfrak{M}_3{}^8)_{\mathfrak{K}}$.

It is clear that a necessary and sufficient condition that a central simple Jordan algebra $\mathfrak{A}$ of degree k be a reduced algebra is that $\mathfrak{A}$ contain k pairwise orthogonal idempotents. Furthermore it is easy to see that if elements of any linear space are pairwise orthogonal idempotents when multiplication is denoted by $a \cdot b$ they remain so when multiplication is defined by (2).

We require the following lemma concerning certain central simple Jordan algebras of linear transformations, namely, the algebras of odd prime degree which have split algebra of type (3) ([3], p. 554). These algebras $\mathfrak{A}$ have degree k and order $2k^2 - k$ over $\mathfrak{F}$, and for each there exists a scalar extension $\mathfrak{K}$ of $\mathfrak{F}$ such that the split algebra $\mathfrak{A}_{\mathfrak{K}}$ consists of all $2k$-rowed matrices $a = aT$ (with elements in $\mathfrak{K}$) defined by $aT = ga'g^{-1}$ for g in (11).

LEMMA. *Let $\mathfrak{F}$ be a nonmodular field and let k be an odd prime. If $\mathfrak{A}$ is a central simple Jordan algebra (of linear transformations) of degree k over $\mathfrak{F}$ having split algebra $\mathfrak{A}_{\mathfrak{K}}$ of type* (3) ([3], p. 554), *then $\mathfrak{A}$ is a reduced algebra.*

For by [7], Theorem 7, there exists a central simple associative algebra $\mathfrak{B}$ over $\mathfrak{F}$ with involution U such that $\mathfrak{B}_{\mathfrak{K}}$ is the total matric algebra of degree $2k$ over $\mathfrak{K}$ and $\mathfrak{A}$ is the set of U-symmetric elements of $\mathfrak{B}$. Hence $\mathfrak{B}$ has degree $2k$ over $\mathfrak{F}$, and $\mathfrak{B} = \mathfrak{B}_k \times \mathfrak{B}_2$ where $\mathfrak{B}_m$ has degree m over $\mathfrak{F}$ $(m = 2, k)$. In each case m is a prime, so $\mathfrak{B}_m$ is either $\mathfrak{M}_m$ or a division algebra $\mathfrak{D}_m$. First

we see that $\mathfrak{B}_k = \mathfrak{M}_k$, for otherwise either $\mathfrak{B} = \mathfrak{D}_k \times \mathfrak{M}_2$ or $\mathfrak{B} = \mathfrak{D}_k \times \mathfrak{D}_2$ and, since $\mathfrak{B}$ has involution U, it would follow in either case that $\mathfrak{D}_k$ also has an involution ([2], p. 156). But this is impossible since $\mathfrak{D}_k$ has exponent k, while an involutorial central simple associative algebra must have exponent either one or two ([2], p. 161).

Hence either $\mathfrak{B} = \mathfrak{M}_k \times \mathfrak{D}_2$ or $\mathfrak{B} = \mathfrak{M}_k \times \mathfrak{M}_2 = \mathfrak{M}_{2k}$. In the latter case, if U were cogredient with (10), then $\mathfrak{A}$ would have a split algebra of type (2) ([3], p. 554) with degree $2k$, a contradiction. Hence, if $\mathfrak{B} = \mathfrak{M}_{2k} = \mathfrak{M}_k \times \mathfrak{M}_2$, the involution U is cogredient with (11) which induces transposition in $\mathfrak{M}_k$. In case $\mathfrak{B} = \mathfrak{M}_k \times \mathfrak{D}_2$, U induces an involution in $\mathfrak{M}_k$ which, since k is odd, is cogredient with (10). In either case we have k pairwise orthogonal idempotents in $\mathfrak{M}_k \subset \mathfrak{B}$ which are U-symmetric and are therefore in $\mathfrak{A}$. By the remark above, $\mathfrak{A}$ is a reduced algebra.

THEOREM 1. *All exceptional central simple Jordan algebras $\mathfrak{A}$ over a nonmodular field $\mathfrak{F}$ are reduced algebras.*[1]

For $\mathfrak{A} \subseteq \mathfrak{A}_{\mathfrak{K}} \subset (\mathfrak{M}_3 \times \mathfrak{D})_{\mathfrak{K}} = \mathfrak{W}_{\mathfrak{K}} + u\mathfrak{W}_{\mathfrak{K}}$ where $u^2 = -1$ and multiplication is given by (12) with $\gamma = -1$. Then every element of $\mathfrak{A}$ (being in $\mathfrak{A}_{\mathfrak{K}}$ and therefore Hermitian) may be written uniquely in the form $a = w_a + uw^*_a$ for Hermitian elements $w_a = \bar{w}_a'$ and $uw^*_a = -uw^*_a{}'$ in $\mathfrak{W}_{\mathfrak{K}} = (\mathfrak{M}_3 \times \mathfrak{Q})_{\mathfrak{K}}$ and $u\mathfrak{W}_{\mathfrak{K}}$ respectively. Let $\mathfrak{S}$ be the set of all w_a and $\mathfrak{T}$ the set of all uw^*_a for a in $\mathfrak{A}$. Then $\mathfrak{S}$ and $\mathfrak{T}$ are linear sets of dimensions s and t respectively over $\mathfrak{F}$ such that $\mathfrak{A} \subseteq \mathfrak{S} + \mathfrak{T}$. Hence $27 \leqq s + t$. Now $\mathfrak{S}_{\mathfrak{K}}$ is contained in the set $\mathfrak{A}_{\mathfrak{K}} \cap \mathfrak{W}_{\mathfrak{K}}$ of all Hermitian elements in $\mathfrak{W}_{\mathfrak{K}}$ (a linear set of order 15 over $\mathfrak{K}$). Hence $s \leqq 15$. Similarly $\mathfrak{T}_{\mathfrak{K}}$ is contained in the set $\mathfrak{A}_{\mathfrak{K}} \cap u\mathfrak{W}_{\mathfrak{K}}$ of all Hermitian elements in $u\mathfrak{W}_{\mathfrak{K}}$ (a linear set of order 12 over $\mathfrak{K}$), so that $t \leqq 12$. Then $\mathfrak{S} + \mathfrak{T}$ has order $s + t = 27$ over $\mathfrak{F}$ and we have $\mathfrak{A} = \mathfrak{S} + \mathfrak{T}$, $s = 15$, $t = 12$. That is, the elements w_a of $\mathfrak{S}$ are actually elements of $\mathfrak{A}$, and also $\mathfrak{S}_{\mathfrak{K}} = \mathfrak{A}_{\mathfrak{K}} \cap \mathfrak{W}_{\mathfrak{K}}$. Then $\mathfrak{S}$ is a Jordan algebra of linear transformations since $\mathfrak{S}^2 = \mathfrak{S}$ and $\mathfrak{W}_{\mathfrak{K}}$ is associative. Since $\mathfrak{S}_{\mathfrak{K}}$ is the set of Hermitian elements $w = \bar{w}'$ of $\mathfrak{W}_{\mathfrak{K}}$, it follows that $\mathfrak{S}$ has split algebra of type (3) ([3], p. 554), the degree being three. By the preceding lemma, $k = 3$, we have $\mathfrak{S}$ a reduced algebra. But then $\mathfrak{A}$, containing $\mathfrak{S}$, contains three pairwise orthogonal idempotents and is itself a reduced algebra.

---

[1] Thanks are due to C. Chevalley for the correction of an error in the original proof of this theorem.

3. Reduced algebras. For any Cayley-Dickson algebra $\mathfrak{C}$ over $\mathfrak{F}$ of characteristic not two, the set $\mathfrak{A}$ of J-symmetric elements of $\mathfrak{M}_3 \times \mathfrak{C}$, with involution J as in (3), is an exceptional central simple Jordan algebra over $\mathfrak{F}$. For $\mathfrak{A}$ has $\mathfrak{M}_3{}^8$ as split algebra. If $\mathfrak{C}$ has norm form $n(x) = xCx'$ with C as in (8), while $p = \operatorname{diag}\{\pi_1, \pi_2, \pi_3\}$, we merely adjoin to $\mathfrak{F}$ the square roots of the scalars $-\gamma_i$ and π_i to obtain the necessary scalar extension $\mathfrak{K}$ over $\mathfrak{F}$ such that $\mathfrak{A}_{\mathfrak{K}} = (\mathfrak{M}_3{}^8)_{\mathfrak{K}}$. For then $H = \operatorname{diag}\{1, \sqrt{-\gamma_1}, \cdots, \sqrt{-\gamma_7}\}$ is the matrix of a linear transformation on $\mathfrak{C}_{\mathfrak{K}}$ such that $1H = 1$, $HH' = C$. Then $d(xH) = (xH)(xH)' = xHH'x' = xCx' = n(x)$ so that $\mathfrak{D}_{\mathfrak{K}}$ and $\mathfrak{C}_{\mathfrak{K}}$ have equivalent norm forms. Then $\mathfrak{C}_{\mathfrak{K}} \cong \mathfrak{D}_{\mathfrak{K}}$ ([5], p. 777). Hence $(\mathfrak{M}_3 \times \mathfrak{C})_{\mathfrak{K}} = \mathfrak{M}_3 \times \mathfrak{C}_{\mathfrak{K}} = \mathfrak{M}_3 \times \mathfrak{D}_{\mathfrak{K}}$ so that $\mathfrak{A}_{\mathfrak{K}}$ consists of the J-symmetric elements of $\mathfrak{M}_3 \times \mathfrak{D}_{\mathfrak{K}}$. Let $s = \operatorname{diag}\{\sqrt{\pi_1}, \sqrt{\pi_2}, \sqrt{\pi_3}\}$. Then $\mathfrak{A}_{\mathfrak{K}}$ and $(\mathfrak{M}_3{}^8)_{\mathfrak{K}}$ are equivalent under the correspondence

$$a \to s^{-1}as. \tag{13}$$

For $s^2 = p$ and $s = \bar{s} = s'$ imply that $(\overline{s^{-1}as})' = s\bar{a}'s^{-1} = sp^{-1}aps^{-1} = s^{-1}as$ if and only if $a = p\bar{a}'p^{-1}$. Hence s^{-1} as is in $(\mathfrak{M}_3{}^8)_{\mathfrak{K}}$ if and only if a is in $\mathfrak{A}_{\mathfrak{K}}$. The correspondence (13) is then obviously an equivalence between $\mathfrak{A}_{\mathfrak{K}}$ and $(\mathfrak{M}_3{}^8)_{\mathfrak{K}}$. Thus any such algebra $\mathfrak{A}$ is an exceptional central simple Jordan algebra over $\mathfrak{F}$.

The J-symmetric elements of $\mathfrak{M}_3 \times \mathfrak{C}$ may be written as matrices (x_{ij}) with $x_{ij} = \pi_i \pi_j{}^{-1} \bar{x}_{ji}$ in $\mathfrak{C}$ $(i, j = 1, 2, 3)$. For $i = j$ we have $x_{ii} = \bar{x}_{ii} = \xi_i$ in $\mathfrak{F}$. Hence any element a of $\mathfrak{A}$ has the form

$$a = \begin{pmatrix} \xi_1 & x_{12} & \pi_1\pi_3{}^{-1}\bar{x}_{31} \\ \pi_2\pi_1{}^{-1}\bar{x}_{12} & \xi_2 & x_{23} \\ x_{31} & \pi_3\pi_2{}^{-1}\bar{x}_{23} & \xi_3 \end{pmatrix} \tag{14}$$

for x_{ij} in $\mathfrak{C}$, $p = \operatorname{diag}\{\pi_1, \pi_2, \pi_3\}$.

Theorem 1 insures that, if $\mathfrak{F}$ is nonmodular, $\mathfrak{A}$ is a reduced algebra. However, to see in general that $\mathfrak{A}$ is a reduced algebra it is easiest simply to pick out the three pairwise orthogonal idempotents e_i of $\mathfrak{A}$ with 1 in the i-th row and i-th column of (14), zeros elsewhere.

We write

$$a = (\xi_1, \xi_2, \xi_3, x_{12}, x_{23}, x_{31}) \tag{15}$$

as a vector of dimension 27 over $\mathfrak{F}$. Then if $b = (\eta_1, \eta_2, \eta_3, y_{12}, y_{23}, y_{31})$, we have $ab = ba = c = (\zeta_1, \zeta_2, \zeta_3, z_{12}, z_{23}, z_{31})$, where, for cyclic permutations (i, j, k) of $(1, 2, 3)$,

$$(16)\qquad \zeta_i = \xi_i\eta_i + \pi_j\pi_i^{-1}t(x_{ij}\bar{y}_{ij}) + \pi_i\pi_k^{-1}t(\bar{x}_{ki}y_{ki}) \text{ in } \mathfrak{F}$$

and

$$(17)\qquad 2z_{ij} = (\xi_i + \xi_j)y_{ij} + (\eta_i + \eta_j)x_{ij} + \pi_i\pi_j^{-1}(\bar{x}_{ki}\bar{y}_{jk} + \bar{y}_{ki}\bar{x}_{jk}) \text{ in } \mathfrak{C}.$$

In the special case $\mathfrak{M}_3{}^8$, we have $t(x\bar{y}) = 2xy' = 2yx'$ and $\pi_i = 1$ $(i = 1, 2, 3)$ so the right multiplication R_a for a in $\mathfrak{M}_3{}^8$ has matrix

$$(18)\qquad \frac{1}{2}\begin{bmatrix} 2\xi_1 & 0 & 0 & x_{12} & 0 & x_{31} \\ 0 & 2\xi_2 & 0 & x_{12} & x_{23} & 0 \\ 0 & 0 & 2\xi_3 & 0 & x_{23} & x_{31} \\ 2x_{12}' & 2x_{12}' & 0 & (\xi_1+\xi_2)I & L_{x_{31}}S & R_{x_{23}}S \\ 0 & 2x_{23}' & 2x_{23}' & R_{x_{31}}S & (\xi_2+\xi_3)I & L_{x_{12}}S \\ 2x_{31}' & 0 & 2x_{31}' & L_{x_{23}}S & R_{x_{12}}S & (\xi_1+\xi_3)I \end{bmatrix},$$

where R_x and L_x are right and left multiplications in $\mathfrak{D}$ and

$$(19)\qquad S = \operatorname{diag}\{1, -I_7\}$$

is the matrix of the involution $x \to \bar{x}$ in $\mathfrak{D}$.

THEOREM 2. *An exceptional central simple Jordan algebra $\mathfrak{A}$ over a nonmodular field $\mathfrak{F}$ is the set of J-symmetric elements of $\mathfrak{M}_3 \times \mathfrak{C}$, where $\mathfrak{C}$ is a Cayley-Dickson algebra over $\mathfrak{F}$ and J is the involution* (3) *of $\mathfrak{M}_3 \times \mathfrak{C}$; multiplication in $\mathfrak{A}$ is defined by* (2). *If $\mathfrak{F}$ has characteristic not two, then a reduced Jordan algebra over $\mathfrak{F}$, having $\mathfrak{M}_3{}^8$ as split algebra, has this form.*

The first statement in the theorem follows directly from the second and Theorem 1. Therefore we consider a Jordan algebra $\mathfrak{A}$ over $\mathfrak{F}$ (of characteristic not two) with three pairwise orthogonal idempotents e_i such that $1 = e_1 + e_2 + e_3$, and a scalar extension $\mathfrak{K}$ of $\mathfrak{F}$ such that $\mathfrak{A}_{\mathfrak{K}} \cong (\mathfrak{M}_3{}^8)_{\mathfrak{K}}$. If we had assumed $\mathfrak{F}$ nonmodular, we could write immediately

$$(20)\qquad \mathfrak{A} = e_1\mathfrak{F} + e_2\mathfrak{F} + e_3\mathfrak{F} + \mathfrak{A}_{12} + \mathfrak{A}_{23} + \mathfrak{A}_{31},\ \mathfrak{A}_{ij} = \mathfrak{A}_{e_i}(\tfrac{1}{2}) \cap \mathfrak{A}_{e_j}(\tfrac{1}{2}),$$

since $\mathfrak{A}$ is a reduced algebra of degree three ([4], § 15). However, one may see from [4], §§ 10, 14, that the decomposition of $\mathfrak{A}$ into the supplementary sum of the spaces $\mathfrak{A}_{ii}$, $\mathfrak{A}_{ij}$ $(i \neq j)$ is valid over any field of characteristic not two. Therefore we have $\mathfrak{A} = \mathfrak{A}_{11} + \mathfrak{A}_{22} + \mathfrak{A}_{33} + \mathfrak{A}_{12} + \mathfrak{A}_{23} + \mathfrak{A}_{31}$. We abbreviate $(\mathfrak{M}_3{}^8)_{\mathfrak{K}}$ as $\mathfrak{M}$, and denote the equivalence between $\mathfrak{A}_{\mathfrak{K}}$ and $\mathfrak{M}$ by P. Now $e_iP = h_i$ for pairwise orthogonal idempotents h_i of $\mathfrak{M}$ satisfying $1 = h_1 + h_2 + h_3$, and $\mathfrak{M} = h_1\mathfrak{K} + h_2\mathfrak{K} + h_3\mathfrak{K} + \mathfrak{M}_{12} + \mathfrak{M}_{23} + \mathfrak{M}_{31}$ for $\mathfrak{M}_{ij} = \mathfrak{M}_{h_i}(\frac{1}{2}) \cap \mathfrak{M}_{h_j}(\frac{1}{2})$. But $(\mathfrak{A}_{ii})_{\mathfrak{K}}P = h_i\mathfrak{K}$ and $(\mathfrak{A}_{ij})_{\mathfrak{K}}P = \mathfrak{M}_{ij}$ so

that $\mathfrak{A}_{ii} = e_i\mathfrak{F}$. That is, (20) holds. Moreover, if we write any element a of $\mathfrak{A}$ in the form (15) corresponding to the decomposition (20) we have the matrix

$$P = \operatorname{diag}\{I_3, P_{12}, P_{23}, P_{31}\} \tag{21}$$

of the equivalence P between $\mathfrak{A}_{\mathfrak{K}}$ and $\mathfrak{M}$, where I_3 is the 3-rowed identity and the P_{ij} are nonsingular 8-by-8 matrices with elements in $\mathfrak{K}$. The right multiplication $R_a^{(0)}$ of a in $\mathfrak{A}$ then has the form

$$R_a^{(0)} = PR_{aP}P^{-1} \tag{22}$$

where R_a is the right multiplication (18) of a in $\mathfrak{M}_3^8$. It is important to note that if a has coefficients in $\mathfrak{F}$, the elements of R_a and $R_a^{(0)}$ are in $\mathfrak{F}$ even though the elements of P are in $\mathfrak{K}$.

We compute $R_a^{(0)}$ from (18), (21), (22). The first three rows are the same as in (18). Below them in the first three columns we must replace $2x_{ij}'$ by

$$2P_{ij}P_{ij}'x_{ij}'. \tag{23}$$

In the remaining 24-by-24 matrix the scalar submatrices are unchanged but we must replace $R_{x_{ij}}S$ by

$$P_{ki}R_{x_{ij}P_{ij}}SP_{jk}^{-1} \tag{24}$$

and $L_{x_{ij}}S$ by

$$P_{jk}L_{x_{ij}P_{ij}}SP_{ki}^{-1} \tag{25}$$

for cyclic permutations (i, j, k) of $(1, 2, 3)$. The elements of the matrices (23), (24), (25) are in $\mathfrak{F}$ for all x_{ij} with coefficients in $\mathfrak{F}$. In particular, it follows that the non-singular symmetric matrices $P_{ij}P_{ij}'$ have elements in $\mathfrak{F}$. Hence there exist changes of basis in $\mathfrak{A}_{ij}$ (over $\mathfrak{F}$) such that

$$P_{ij}P_{ij}' = C_{ij} \tag{26}$$

are non-singular diagonal matrices with elements in $\mathfrak{F}$. We assume that such a change of basis has been made. We write e for the 1-by-8 vector $(1, 0, \cdots, 0)$ in $\mathfrak{A}_{ij}$. Then $p_{ij} = eP_{ij}$ may be regarded as elements of $\mathfrak{D}_{\mathfrak{K}}$, and we have $d(p_{ij}) = (eP_{ij})(eP_{ij})' = eC_{ij}e' \neq 0$ in $\mathfrak{F}$. We define, for cyclic permutations (i, j, k) of $(1, 2, 3)$.

$$d(p_{ij}) = \pi_k \neq 0 \text{ in } \mathfrak{F}. \tag{27}$$

It follows from (24) that the vector $h = eP_{23}R_{eP_{31}}SP_{12}^{-1} = (p_{23}p_{31})SP_{12}^{-1}$ has coefficients in $\mathfrak{F}$. Hence by (24) and (27) the matrix

$$H = \pi_2^{-1} P_{31} R_{hP_{12}} S P_{23}^{-1} = \pi_2^{-1} P_{31} S L_{p_{23}p_{31}} P_{23}^{-1}$$

has elements in $\mathfrak{F}$. Moreover,

$$eH = \pi_2^{-1}\{(p_{23}p_{31})\bar{p}_{31}\} P_{23}^{-1} = p_{23} P_{23}^{-1} = e. \tag{28}$$

By (25) the matrix

$$G = P_{12} L_{eP_{31}} S P_{23}^{-1} = P_{12} L_{p_{31}} S P_{23}^{-1}$$

also has elements in $\mathfrak{F}$. Write $Q_{12} = G^{-1}P_{12}$, $Q_{23} = P_{23}$, $Q_{31} = H^{-1}P_{31}$. Then

$$Q_{12} L_{eQ_{31}} = Q_{23} S = Q_{31} R_{eQ_{12}}. \tag{29}$$

For $G^{-1}P_{12}L_{eH^{-1}P_{31}} = P_{23}SL_{p_{31}}{}^{-1}L_{eP_{31}} = P_{23}S$ by (28). Since $eG^{-1}P_{12} = \bar{p}_{23}L_{p_{31}}{}^{-1} = \pi_2^{-1}\bar{p}_{31}\bar{p}_{23}$ we have also $H^{-1}P_{31}R_{eG^{-1}P_{12}} = P_{23}L_{p_{23}p_{31}}{}^{-1}SR_{\bar{p}_{31}\bar{p}_{23}} = P_{23}S$.

Since G and H have their elements in $\mathfrak{F}$ we are able to effect changes of basis (over $\mathfrak{F}$) in $\mathfrak{A}_{12}$ with matrix G^{-1} and in $\mathfrak{A}_{31}$ with matrix H^{-1}. The right multiplication $R_a{}^{(0)}$ has the same matrix as before except that P_{ij} has been replaced throughout by Q_{ij}. Now we change notation, writing P_{ij} for Q_{ij}, so that $R_a{}^{(0)}$ has its original matrix, while (29) yields

$$P_{12} L_{p_{31}} = P_{23} S = P_{31} R_{p_{12}}. \tag{30}$$

This change has not affected P_{23} so $P_{23}P_{23}' = C_{23}$ in (26) is still a diagonal matrix with elements in $\mathfrak{F}$. From (30) we have

$$p_{31} \cdot xP_{12} = \overline{xP_{23}} = xP_{31} \cdot p_{12} \tag{31}$$

for all x in $\mathfrak{D}$. Since $d(x)$ permits composition, this yields $\pi_2 d(xP_{12}) = d(xP_{23}) = \pi_3 d(xP_{31})$. Moreover,

$$p_{31}p_{12} = \bar{p}_{23} \tag{32}$$

also follows from (31). Hence

$$\pi_2\pi_3 = \pi_1 \tag{33}$$

and

$$\pi_3^{-1} d(xP_{12}) = \pi_1^{-1} d(xP_{23}) = \pi_2^{-1} d(xP_{31}) \tag{34}$$

for all x in $\mathfrak{D}$. Equivalently, $\pi_3^{-1}P_{12}P_{12}' = \pi_1^{-1}P_{23}P_{23}' = \pi_2^{-1}P_{31}P_{31}'$ since the matrices $\pi_k^{-1}P_{ij}P_{ij}'$ of the quadratic forms (34) are symmetric. Since the matrix $P_{23}P_{23}' = C_{23}$ is a nonsingular diagonal matrix with elements in $\mathfrak{F}$, so is

$$C = \pi_k^{-1} P_{ij} P_{ij}' = \pi_k^{-1} C_{ij} \tag{35}$$

for cyclic permutations (i, j, k) of $(1, 2, 3)$.

Since the matrices (25) have elements in $\mathfrak{F}$ for all x with coordinates in $\mathfrak{F}$, we may define a (non-associative) algebra $\mathfrak{C}$ of order 8 over $\mathfrak{F}$ by writing multiplication $y \circ x = yR_x^{(0)}$ in $\mathfrak{C}$ where

$$(36) \qquad R_x^{(0)} = P_{12}L_{xP_{31}}SP_{23}^{-1} = P_{12}SR_{xP_{31}S}P_{23}^{-1}.$$

Then the left multiplication $L_x^{(0)}$ in $\mathfrak{C}$ is

$$(37) \qquad L_x^{(0)} = P_{31}SL_{xP_{12}S}P_{23}^{-1}.$$

Now e is the unity quantity of $\mathfrak{C}$ since $R_e^{(0)} = P_{12}L_{p_{31}}SP_{23}^{-1} = I$ and $L_e^{(0)} = P_{31}SL_{\bar{p}_{12}}P_{23}^{-1} = P_{31}^{-1}R_{p_{12}}SP_{23}^{-1} = I$ by (30). Moreover, $\mathfrak{C}$ is isotopic (over $\mathfrak{K}$) to $\mathfrak{D}$. Hence $\mathfrak{C}_{\mathfrak{K}} \cong \mathfrak{D}_{\mathfrak{K}}$ and $\mathfrak{C}$ is a Cayley-Dickson algebra over $\mathfrak{F}$ ([8], Theorem 4). Then $\mathfrak{C}$ has a norm form $n(x)$ satisfying

$$\begin{aligned}\{n(x)\}^4 = |R_x^{(0)}| &= |P_{12}L_{xP_{31}}SP_{23}^{-1}| = |P_{12}| \cdot |P_{23}|^{-1} \cdot |L_{xP_{31}}| \\ &= |L_{p_{31}}|^{-1} \cdot |L_{xP_{31}}| = \{\pi_2^{-1}d(xP_{31})\}^4\end{aligned}$$

by (30). Hence $n(x) = \epsilon\pi_2^{-1}d(xP_{31})$ where $\epsilon^4 = 1$. For $x = e$ we have $1 = n(e) = \epsilon$, so $n(x)$ equals the common value of (34). That is, $n(x) = xCx'$ for C in (35). Also $\mathfrak{C}$ has the involution S_0 satisfying $xS_0 \circ x = n(x)e$ by (5). Hence

$$\begin{aligned}xS_0 = n(x)eR_x^{(0)-1} &= n(x)p_{23}R_{xP_{31}S}^{-1}SP_{12}^{-1} = \pi_2^{-1}p_{23}R_{xP_{31}}SP_{12}^{-1} \\ &= \pi_2^{-1}(p_{23} \cdot xP_{31})SP_{12}^{-1} = \pi_2^{-1}xP_{31}L_{p_{23}}SP_{12}^{-1}\end{aligned}$$

so that

$$(38) \qquad S_0 = \pi_2^{-1}P_{31}L_{p_{23}}SP_{12}^{-1}.$$

Since $S_0 = S_0^{-1}$ we have also $S_0 = \pi_3^{-1}P_{12}R_{p_{23}}SP_{31}^{-1}$.

We write $T_{12} = \pi_3^{-1}S_0P_{12} = \pi_1^{-1}P_{31}L_{p_{23}}S$, $T_{23} = \pi_2^{-1}P_{23}$, $T_{31} = S_0P_{31} = \pi_3^{-1}P_{12}R_{p_{23}}S$. Using (9), it is easy to verify that

$$(39) \qquad T_{ij}T_{ij}' = \pi_j\pi_i^{-1}C$$

for C in (35). Also

$$(40) \qquad T_{ki}R_{xT_{ij}}ST_{jk}^{-1} = \pi_j\pi_k^{-1}R_x^{(0)}S_0$$

and

$$(41) \qquad T_{jk}L_{xT_{ij}}ST_{ki}^{-1} = \pi_k\pi_i^{-1}L_x^{(0)}S_0$$

for cyclic permutations (i, j, k) of $(1, 2, 3)$. Verification of the six equations (40), (41) depends only on the verification of one equation in each of the corresponding pairs, for the products on the left and right sides of corresponding pairs are equal. That is,

$$(T_{ki}R_{xT_{ij}}ST_{jk}^{-1})(T_{jk}L_{xT_{ij}}ST_{ki}^{-1}) = d(xT_{ij})I = \pi_j\pi_i^{-1}n(x)I$$

since

$$d(xT_{ij}) = xT_{ij}T_{ij}'x' = \pi_j\pi_i^{-1}xCx' = \pi_j\pi_i^{-1}n(x)$$

by (39), while

$$(\pi_j\pi_k^{-1}R_x^{(0)}S_0)(\pi_k\pi_i^{-1}L_x^{(0)}S_0) = \pi_j\pi_i^{-1}n(x)I$$

also. The proof of (41) for $(i, j, k) = (3, 1, 2)$ follows directly from the definition (36):

$$T_{12}L_{xT_{31}}ST_{23}^{-1} = \pi_2\pi_3^{-1}S_0P_{12}L_{xS_0P_{31}}SP_{23}^{-1} = \pi_2\pi_3^{-1}S_0R_{xS_0}^{(0)} = \pi_2\pi_3^{-1}L_x^{(0)}S_0,$$

and equation (40) for $(i, j, k) = (1, 2, 3)$ similarly from (37). To prove equation (40) for $(i, j, k) = (2, 3, 1)$ we need to apply twice a lemma of Moufang ([8], Lemma 1). First we note that $P_{12}S = \pi_2^{-1}S_0P_{31}L_{p_{23}}$ by (38), so that by (30) and (32) we have

$$P_{12}S = \pi_2^{-1}S_0P_{12}L_{p_{31}}R^{-1}{}_{p_{12}}L_{\bar{p}_{12}\bar{p}_{31}} = \pi_1^{-1}S_0P_{12}L_{p_{31}}(R_{\bar{p}_{12}}L_{\bar{p}_{12}\bar{p}_{31}})$$
$$= \pi_1^{-1}S_0P_{12}L_{p_{31}}(L_{\bar{p}_{31}}L_{\bar{p}_{12}}R_{\bar{p}_{12}}) = \pi_3^{-1}S_0P_{12}L_{\bar{p}_{12}}R_{\bar{p}_{12}} = T_{12}L_{\bar{p}_{12}}R_{\bar{p}_{12}}.$$

Using this, we have

$$\pi_3\pi_1^{-1}R_x^{(0)}S_0 = \pi_2^{-1}P_{12}SR_{xP_{31}S}P_{23}^{-1}S_0 = \pi_2^{-1}T_{12}L_{\bar{p}_{12}}R_{\bar{p}_{12}}R_{xP_{31}S}SR_{p_{12}}^{-1}P_{31}^{-1}S_0$$
$$= \pi_2^{-1}T_{12}(L_{\bar{p}_{12}}R_{\bar{p}_{12}}R_{xP_{31}S}L_{\bar{p}_{12}}^{-1})SP_{31}^{-1}S_0 = \pi_2^{-1}T_{12}(R_{\bar{p}_{12}\cdot xP_{31}S})ST_{31}^{-1}$$
$$= \pi_2^{-1}T_{12}R_{xP_{23}}ST_{31}^{-1}$$

by (31), or $\pi_3\pi_1^{-1}R_x^{(0)}S_0 = T_{12}R_{xT_{23}}ST_{31}^{-1}$ as desired.

We again change basis in $\mathfrak{A}_{ij}$ (over $\mathfrak{F}$), this time with matrix $\pi_3^{-1}S_0$ in $\mathfrak{A}_{12}$, $\pi_2^{-1}I$ in $\mathfrak{A}_{23}$, and S_0 in $\mathfrak{A}_{31}$. The right multiplication $R_a^{(0)}$ of the element a in $\mathfrak{A}$ has the same matrix as before except that P_{ij} has been replaced throughout by T_{ij}. That is, the first three rows of $R_a^{(0)}$ are the same as in (18). A comparison of (23), (24), (25) with (39), (40), (41) shows that below the first three rows we have in the first three columns substituted $2\pi_j\pi_i^{-1}Cx_{ij}'$ for the $2x_{ij}'$ in (18), while in the 24-by 24 matrix remaining the scalar submatrices are unchanged but we have $\pi_j\pi_k^{-1}R_{x_{ij}}^{(0)}S_0$ instead of $R_{x_{ij}}S$ and $\pi_k\pi_i^{-1}L_{x_{ij}}^{(0)}S_0$ instead of $L_{x_{ij}}S$. But then $R_a^{(0)}$ has exactly the matrix of the right multiplication yielding (16) and (17), where $\mathfrak{C}$ is the Cayley-Dickson algebra with multiplication $y \circ x = yR_x^{(0)}$, involution $x \to \bar{x} = xS_0$, and associated symmetric bilinear form $\frac{1}{2}t(x \circ \bar{y}) = xCy'$ corresponding to $n(x) = xCx'$. Our theorem is proved.

4. Inequivalent exceptional Jordan algebras. The only exceptional central simple Jordan algebra over an algebraically closed nonmodular field is $\mathfrak{M}_3^8$. In order to show that the algebras defined in **3** are a valid generalization of $\mathfrak{M}_3^8$ we ought to exhibit fields $\mathfrak{F}$ over which there exist inequivalent algebras of that type.

Let the base field $\mathfrak{F}$ be formally real; then there exist inequivalent Jordan algebras over $\mathfrak{F}$ with $\mathfrak{M}_3^8$ as split algebra. For then $\mathfrak{D}$ is a division algebra over $\mathfrak{F}$ and cannot be equivalent to the (unique) Cayley-Dickson algebra over $\mathfrak{F}$ which has divisors of zero; the existence of inequivalent central simple Jordan algebras over $\mathfrak{F}$ then follows from the following sharper result.

THEOREM 3. *Let $\mathfrak{C}$ and $\mathfrak{C}^*$ be inequivalent Cayley-Dickson algebras over $\mathfrak{F}$ of characteristic not two. Then the exceptional central simple Jordan algebras $\mathfrak{A}$ and $\mathfrak{A}^*$ consisting of the 3-rowed Hermitian matrices with elements in $\mathfrak{C}$ and $\mathfrak{C}^*$ respectively, multiplication being defined by* (2), *are inequivalent over $\mathfrak{F}$.*

Since there is (in the sense of equivalence) only one Cayley-Dickson algebra over $\mathfrak{F}$ which has divisors of zero, at least one of the algebras $\mathfrak{C}$, $\mathfrak{C}^*$ is a division algebra over $\mathfrak{F}$; say it is $\mathfrak{C}$. We take normalized bases in $\mathfrak{C}$ and $\mathfrak{C}^*$ so that each has involution (19). The associated symmetric bilinear forms are $\frac{1}{2}t(x\bar{y})$ and $\frac{1}{2}t(x * \bar{y})$ respectively and the corresponding norm forms are

$$n(x) = \tfrac{1}{2}t(x\bar{x}) \text{ and } n^*(x) = \tfrac{1}{2}t(x * \bar{x}). \tag{42}$$

Writing the general element a of $\mathfrak{A}$ and $\mathfrak{A}^*$ in the form (15), multiplication in $\mathfrak{A}$ is given by (16) and (17) with $\pi_i = 1$ $(i = 1, 2, 3)$, and in $\mathfrak{A}^*$ by $a * b = b * a = c^* = (\zeta^*_1, \zeta^*_2, \zeta^*_3, z^*_{12}, z^*_{23}, z^*_{31})$ where we have, in particular,

$$\zeta^*_1 = \xi_1\eta_1 + t(x_{12} * \bar{y}_{12}) + t(\bar{x}_{31} * y_{31}). \tag{43}$$

Our proof is indirect. If $\mathfrak{A}$ and $\mathfrak{A}^*$ are equivalent, the equivalence P has matrix (21) for non-singular P_{ij} with elements in $\mathfrak{F}$. Then $c^*P = (a * b)P = aP \cdot bP$ implies, in particular,

$$\zeta^*_1 = \xi_1\eta_1 + t(x_{12}P_{12} \cdot \overline{y_{12}P_{12}}) + t(\overline{x_{31}P_{31}} \cdot y_{31}P_{31}). \tag{44}$$

Let $y_{12} = x_{12}$, $y_{31} = 0$. Then by (42), (43), (44) we have $\xi_1\eta_1 + 2n^*(x_{12}) = \xi_1\eta_1 + 2n(x_{12}P_{12})$ or

$$n^*(x) = n(xP_{12}) \tag{45}$$

for all x in $\mathfrak{C}^*$. Now $1P_{12} = p_{12} \neq 0$ in $\mathfrak{C}$ since P_{12} is non-singular. Since $\mathfrak{C}$ is a division algebra, we have $L_{p_{12}}$ non-singular, and may write $H = P_{12}L_{p_{12}}^{-1}$. Then $1H = 1P_{12}L_{p_{12}}^{-1} = p_{12}L_{p_{12}}^{-1} = 1$, and by (45) we have

$$n^*(x) = n(xHL_{p_{12}}) = n(p_{12} \cdot xH) = n(p_{12})n(xH) = n(xH)$$

since $n(p_{12}) = n(1P_{12}) = n^*(1) = 1$. Hence $n^*(x)$ and $n(x)$ are equivalent norm forms, and $\mathfrak{C} \cong \mathfrak{C}^*$ ([5], p. 777), a contradiction.

Remark. We have shown incidentally the existence of an exceptional central simple Jordan algebra over the field $\mathfrak{R}$ of all real numbers which is different from $\mathfrak{M}_3{}^8$. Of course this algebra is not an "r-number system" ([6]); it fails to be a "formally real" algebra, and actually contains elements whose squares are zero.

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, NEW JERSEY.

REFERENCES

1. A. A. Albert, "On a certain algebra of quantum mechanics," *Annals of Mathematics*, vol. 35 (1934), pp. 65-73.
2. ———, *Structure of algebras*, American Mathematical Society Colloquium Publications, vol. 24 (1939).
3. ———, "On Jordan algebras of linear transformations," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 524-555.
4. ———, "A structure theory for Jordan algebras," *Annals of Mathematics*, vol. 48 (1947), pp. 546-567.
5. N. Jacobson, "Cayley numbers and normal simple Lie algebras of type G," *Duke Mathematical Journal*, vol. 5 (1939), pp. 775-783.
6. P. Jordan, J. v. Neumann, and E. Wigner, "On an algebraic generalization of the quantum mechanical formalism," *Annals of Mathematics*, vol. 35 (1934), pp. 29-64.
7. G. K. Kalisch, "On special Jordan algebras," *Transactions of the American Mathematical Society*, vol. 61 (1947), pp. 482-494.
8. R. D. Schafer, "Alternative algebras over an arbitrary field," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 549-555.

SOLUTION OF A PROBLEM OF R. L. WILDER.*

By R. H. Bing.

The following problem was proposed by R. L. Wilder in "Concerning Simple Continuous Curves and Related Point Sets," *American Journal of Mathematics*, vol. 53 (1931), pp. 39-55: *Suppose that M is a connected im kleinen set which is the sum of two sets M_1 and M_2 such that each is irreducibly connected from the point A to the point B and such that the common part of M_1 and M_2 is A and B. Is M necessarily a simple closed curve?* Wilder showed that M is a simple closed curve if it is a locally compact continuum. This note gives an example to show that M need not be a simple closed curve if this extra requirement is not placed on it.

We shall prove the following for the plane: *Suppose that S is a square plus its interior. There exists a collection G of point sets filling up S such that each element of G is irreducibly connected from a point A to a point B, the common part of two elements of G is $A + B$, and the sum of two elements of G is a locally connected subset of S which is dense in S.*

Let A and B be opposite vertices of S and let E and F be the other vertices. Denote by $[X]$ the collection of all straight line intervals EP where P is a point of $AF + FB - (A + B)$. For each element X of $[X]$, let $L(X)$ denote the collection of all straight line intervals which have their end points on the boundary of S and which are parallel to X. We note that with respect to its elements, each $L(X)$ is an open arc from A to B. Let $[L]$ be the set of all such collections $L(X)$.

There exists a well ordered sequence $W(L)$ such that the elements of $W(L)$ are the elements of $[L]$ and such that if L is an element of $[L]$, then the set of all elements of $[L]$ which precede L in $W(L)$ has a power less than that of $[L]$. We note that $[L]$ has the power of the continuum. We shall denote the element of $W(L)$ whose ordinal number is α by L_α.

Let $W(S)$ be a well ordering of the points of S.

Denote the collection of all subcontinua of S by $[C]$. Points as well as nondegenerate continua are included in $[C]$. Let $W(C)$ be a well ordered sequence such that the elements of $W(C)$ are the elements of $[C]$ and such that if C is a subcontinuum of S, then the collection of all elements of $[C]$

* Received March 25, 1947; Presented to the American Mathematical Society, November 30, 1946.

that precede C in $W(C)$ does not have the power of $[C]$. Since $[C]$ has the power of the continuum, this subcollection has a power less than that of the continuum. We shall denote the element of $W(C)$ whose ordinal number is α by C_α.

Before beginning the assigning of points of S to the elements of G, we shall briefly outline what we intend to do. We shall describe collections $M_1, M_2, \cdots, M_\alpha, \cdots$ such that each of these collections is irreducibly connected from A to B. These collections will be the elements of G. To insure that M_β $(\beta = 1, 2, \cdots, \alpha, \cdots)$ will be irreducible from A to B, it will be defined so as to intersect every element of $[C]$ which is not a subset of an element of L_β and so as to intersect each element of L_β in only one point.

Our procedure will be to sort out the points of S into the collections $M_1, M_2, \cdots, M_\alpha, \cdots$. With the exception of some skips, the points of S are sorted out to the elements of G in the following order. First, a point of C_1 is assigned to M_1; next, a point of C_2 is assigned to M_1 and another point of C_2 is assigned to M_2; $\cdots$; then a point of C_α is assigned to M_1, another point of C_α is assigned to $M_2, \cdots$, and another point of C_α is assigned to M_α; $\cdots$. Since only one point of S is assigned at each step, this assigning will require uncountably many steps.

Now for the details of the description of the elements of G. The points A and B are assigned to each of the collections of G. No other point is assigned to any two elements of G.

First, consider C_1. If it intersects $A + B$, no point of it is assigned to an element of G at this stage. If it does not intersect $A + B$, the first point of $W(S)$ in C_1 is assigned to M_1.

Next, consider C_2. No point of it is assigned to M_1 at this stage if either C_2 contains a point already assigned to M_1 or C_2 is a subset of an element of L_1 that contains a previously assigned point of M_1. However, if C_2 is not such a set, then to M_1 is assigned the first point of $W(S)$ that is in the common part of C_2 and the sum of the elements of L_1 containing no previously assigned point of M_1.

Now, if C_2 intersects neither A nor B and is not a point previously assigned to M_1, then to M_2 is assigned the first point of $W(S)$ in the subset of C_2 that has not been assigned previously to M_1. Otherwise, no point of C_2 is assigned to M_2 at this stage.

This process of assigning points of S to elements of G is continued as follows. Suppose that at some stage, U is the collection of points of S that have not been assigned previously to elements of G. Let V be the collection of all elements C_α of $W(C)$ for which there is an element C_β of $W(C)$ such

that C_β is equal to C_α or precedes C_α in $W(C)$, no point of C_α has been assigned to M_β, and C_α is not a subset of an element of L_β that contains a point which has been assigned to M_β. If C_α is a point of U, it is an element of V since no point would be assigned to M_α before C_α were assigned to an element of G. For convenience, suppose that C_α is the first element of V in $W(C)$. Then a point of C_α is assigned to an element of G at this stage as is described below.

Let R be the collection of all elements L_β of $W(L)$ such that no point of C_α has been assigned to M_β and C_α is not a subset of an element of L_β which contains a point which has been assigned to M_β. Suppose L_β is the first element of R in $W(L)$. Then to M_β is assigned the first point of $W(S)$ in the common part of $C_\alpha \cdot U$ and the sum of the elements of L_β containing no previously assigned points of M_β. If β is the same ordinal number as α, this is the first point to be assigned to M_α.

Suppose that at some stage, Y is the set of points of S that have been assigned previously to elements of G in this process and P is a point of $S - Y$. Then P is an element C_γ of $W(C)$. Now Y is not of the power of the continuum since none of the sets $M_1 \cdot Y, M_2 \cdot Y, \cdots, M_\gamma \cdot Y$ has the power of the continuum.

Each point of S was assigned to some element of G in this process for suppose that U is the collection of all points of S that were not assigned to G even though the process were continued. But as is seen above, a point of U is assigned to an element of G.

No point of $S - (A + B)$ was assigned to two elements of G. Hence, S is the sum of the elements of G and the common part of two elements of G is $A + B$.

We shall now show that each element E of L_α contains a point of M_α. Since E contains a collection of nondegenerate subcontinua such that this collection is of the power of the continuum, there is a nondegenerate subcontinuum C_γ of E such that C_γ follows C_α in $W(C)$. Each point of $S - (A + B)$ was assigned to an element of G while considering an element of $W(C)$. Let Z be the set of all points of S that are assigned to elements of G before considering any element of $W(C)$ that follows C_γ in $W(C)$. Since none of the collections $M_1 \cdot Z, M_2 \cdot Z, \cdots, M_\gamma \cdot Z$ has the power of the continuum, Z does not. Then at some stage, an element of $W(C)$ was considered that followed C_γ in $W(C)$. But this element would not have been considered unless a point of the element of L_α containing C_γ had been assigned to M_α. Hence, each element of L_α contains a point of M_α. However, no two points of an element of L_α are assigned to any element of M_α.

Each subcontinuum C of S which is not a subset of any element of L_α contains a point of M_α because C contains a collection R of subcontinua no one of which is a subset of an element of L_α and such that the collection R has the power of the continuum. Hence, some element of R follows C_α in $W(C)$ and a point of it is assigned to M_α.

We have shown that M_α contains a point of each element of L_α and that each subcontinuum of S which is not a subset of an element of L_α contains a point of M_α. Hence, M_α is connected. If it were the sum of two mutually separated sets, some continuum in its complement and in S would separate some two points of it in S. But each subcontinuum of S that separates S is either an element of L_α or intersects two elements of L_α. Since M_α contains only one point of each element of L_α and since with respect to its elements, L_α is an open arc from A to B, M_γ is irreducibly connected from A to B.

We shall show that if M_α and M_β are two elements of G, then $M_\alpha + M_\beta$ is locally connected. Each nondegenerate subcontinuum of S intersects either two elements of L_α or two elements of L_β and hence contains a point of $M_\alpha + M_\beta$. If D is a domain such that $D \cdot S$ is connected, then $D \cdot (M_\alpha + M_\beta)$ is connected, for if it were the sum of two mutually separated sets, then some continuum R in the complement of $D \cdot (M_\alpha + M_\beta)$ would separate some two points of $D \cdot (M_\alpha + M_\beta)$ from each other in the plane. But some nondegenerate subcontinuum of R would be a subset of $D \cdot S$ and would contain a point of $D \cdot (M_\alpha + M_\beta)$. Hence, $M_\alpha + M_\beta$ is locally connected. Not only is it locally connected but it has the additional property that if P is a point, then $M_\alpha + M_\beta + P$ is locally connected. Hence, the sum of two elements of G has property S.

Question. It can be shown that the sum of no two elements of G is an inner limiting (G_δ) set. It would be interesting to know whether or not it can be concluded that M is a simple closed curve if it is a connected im kleinen inner limiting set which is the sum of two sets M_1 and M_2 such that each is irreducibly connected from the point A to the point B and such that the common part of M_1 and M_2 is A and B.

UNIVERSITY OF TEXAS.

A THEORY OF NORMALITY FOR QUASIGROUPS.*

By Fred Kiokemeister.

Let the quasigroup G be homomorphic to the quasigroup G'. If H is the set of antecedents in G of an element in G' and if H is itself a quasigroup, then H is a normal divisor of G defined by the given homomorphism. If G is a loop, then H is the kernel of the homomorphism. In this case H always exists and is uniquely defined. For a quasigroup the situation in general is not as simple. A homomorphism may define no normal divisor, or a homomorphism may define more than one normal divisor.[1] Thus there cannot be constructed for quasigroups a normality theory on the concept of normal divisor or kernel. We have chosen to employ here the equivalence relation which is defined by each homomorphism.[2] These relations are shown to form a modular lattice. In the event that normal divisors exist, structural uniqueness is obtained by the device of restricting attention to the set of normal divisors containing a fixed element.

1. Normal relations. A multiplicative system is a non-vacuous set G of elements $a, b, c, \cdots$ such that for each pair a, b there is defined an ordered product ab which is a uniquely determined element of G. If, further, to each pair a, b there correspond unique elements x and y in G such that $ax = b$, and $ya = b$, then G is a quasigroup. The following lemmas are immediate consequences of the above definitions.

Lemma 1.1. *Any homomorphic image of a quasigroup is a multiplicative system in which the equations $ax = b$ and $ya = b$ have solutions for every pair a, b.*

Lemma 1.2. *A multiplicative system in which the equations $ax = b$ and*

* Received April 9, 1947.

[1] If each element is idempotent, then each element constitutes a normal divisor of G defined by an isomorphism of G on itself, cf. Bruck [6], p. 34. That such quasigroups may be simple has been pointed out by Bruck in [5], p. 169.

[2] Other treatments are obviously possible. For quasigroups see Bruck [5], pp. 166-169 and Garrison [9] and [10], in particular section 2 of [10]. For loops there are Albert [1, 2], Baer [3], and Bruck [6, 7, 8]. The lattice-theoretic viewpoint has been employed by Smiley [15], and in restricted cases by Hausman and Ore [11] and Murdoch [12].

$ya = b$ have solutions for every pair a, b is a quasigroup if and only if $ca = cb$ and $ac = bc$ each implies $a = b$.

Every homomorphism of a quasigroup G on a quasigroup G' induces an equivalence relation β on the elements of G. If a' and b' are respectively the images of a and b, then $a \beta b$ if and only if $a' = b'$. This equivalence relation satisfies the conditions:

i) $ca \beta cb$ implies $a \beta b$,

ii) $ac \beta bc$ implies $a \beta b$,

iii) $a \beta b$ and $c \beta d$ implies $ac \beta bd$.

Condition iii) identifies β as a congruence relation and follows from the homomorphism of G on G'. Since G' is a quasigroup, Lemma 1.2 implies i) and ii).

Conversely every congruence relation induces a homomorphism of G on the multiplicative system G' of equivalence classes under β. Lemmas 1.1 and 1.2 guarantee that if β satisfies i) and ii) G' will be a quasigroup.

An equivalence relation β satisfying i), ii), and iii) will be called a normal relation on G.

Let α be a second normal relation. We shall say that β contains α (and write $\beta \supseteq \alpha$) if and only if $a \alpha b$ implies $a \beta b$. The intersection, $\beta \cap \alpha$, of β and α is defined to be the logical intersection of the relations: $a(\beta \cap \alpha)b$ if and only if $a \beta b$ and $a \alpha b$. The union, $(\beta \cup \alpha)$, of β and α is defined as follows: $a(\beta \cup \alpha)b$ if and only if there exist elements $a_1, a_2, \cdots, a_n$ in G such that $a \beta a_1 \alpha a_2 \beta \cdots a_n \alpha b$. It is easily verified that the union and intersection of normal relation are normal relations, and that $(\beta \cap \alpha)$ and $(\beta \cup \alpha)$ are a greatest lower bound and a least upper bound respectively for β and α.[3]

THEOREM 1.1. *The normal relations on G form a lattice L.*

We shall need the following

LEMMA 1.3. *If β is a normal relation on G and G contains the quasigroup H, then β is a normal relation on H.*

If a, b, c, d are elements of H, conditions i), ii), and iii) are satisfied in G and therefore in H.

Let $R(a, \beta)$ designate the set of elements g of G such that $g \beta a$.

[3] Ore has made these definitions for equivalence relations in [13].

LEMMA 1.4. *If α and β are normal relations and if a is any element of G, then $R(a,\alpha)=R(a,\beta)$ if and only if $\alpha=\beta$, and $R(a,\alpha)\supseteq R(a,\beta)$ if and only if $\alpha\supseteq\beta$.*

Let b and c be any elements of G, and let $b\,\beta\,c$. There exist in G elements x and y such that $c=xa$ and $b=xy$. By i) $xy\,\beta\,xa$ implies $y\,\beta\,a$. If $R(a,\alpha)\supseteq R(a,\beta)$, then $y\,\alpha\,a$, and by iii) $xy\,\alpha\,xa$. Thus $b\,\alpha\,c$, and $\beta\subseteq\alpha$. Conversely, if $\alpha\supseteq\beta$, $g\,\beta\,a$ implies $g\,\alpha\,a$, and $R(a,\alpha)\supseteq R(a,\beta)$. The first part of the lemma follows from the second.

If S and T are sets of elements in G, then ST is the set of elements st where s lies in S and t lies in T; and $S\cap T$ is the set of elements common to S and T.

LEMMA 1.5. *Let α and β be normal relations. If $g\,\alpha\,f$ or $g\,\beta\,f$ where f is an element of $R(a,\alpha)R(b,\beta)$, then g is an element of $R(a,\alpha)R(b,\beta)$.*

Let $g\,\alpha\,cd$ where $c\,\alpha\,a$ and $d\,\beta\,b$, and let $g=xd$. Then $xd\,\alpha\,cd$, and $x\,\alpha\,c$, i.e., $g=xd$ where x lies in $R(a,\alpha)$ and d lies in $R(b,\beta)$. We may use a similar proof if $g\,\beta\,cd$.

THEOREM 1.2. *If α and β are any normal relations on G, and if a and b are any elements of G, then*

$$R(ab,\alpha\cup\beta)=R(a,\alpha)R(b,\beta),[4]$$

and

$$R(a,\alpha\cap\beta)=R(a,\alpha)\cap R(a,\beta).$$

We shall show that if $g(\alpha\cup\beta)ab$, then g is an element of $R(a,\alpha)R(b,\beta)$. The relation $g(\alpha\cup\beta)ab$ implies the existence of a chain $g=g_0\,\alpha\,g_1\,\beta\,g_2\,\alpha\cdots g_n\,\beta\,ab$. Since ab is an element of $R(a,\alpha)R(b,\beta)$, Lemma 1.5 establishes that g_i is an element of the same set for $i=n,n-1,\cdots,0$.

If, on the other hand, $g=cd$ where $c\,\alpha\,a$ and $d\,\beta\,b$, then $cd\,\alpha\,ad$, and $ad\,\beta\,ab$ by iii). This chain implies that $g(\alpha\cup\beta)ab$, and $R(a,\alpha)R(b,\beta)\subseteq R(ab,\alpha\cup\beta)$.

The second part of the theorem is obvious.

THEOREM 1.3. *The lattice L of normal relations on G is modular.*

For any lattice $\alpha\supseteq\gamma$ implies that $\alpha\cap(\beta\cup\gamma)\supseteq(\alpha\cap\beta)\cup\gamma$. It is sufficient to prove that the inclusion is reversible.[5] Let a be any element of G. By Theorem 1.2

[4] Garrison has established this relation for $a=b$ when G is finite in Theorem 2.4 of [10]. In the terminology of Garrison $R(a,\alpha)$ is an invariant complex when α is a normal relation.

[5] See Birkhoff [4], p. 22 and p. 34.

$$R[a^2, \alpha \cap (\beta \cup \gamma)] = R(a^2, \alpha) \cap R(a, \beta)R(a, \gamma),$$

and

$$R[a^2, (\alpha \cap \beta) \cup \gamma] = [R(a, \alpha) \cap R(a, \beta)]R(a, \gamma).$$

Let g be an element of $R(a^2, \alpha) \cap R(a, \beta)R(a, \gamma)$. Then $g = bc$ where $b \, \beta \, a$, $c \, \gamma \, a$, and $bc \, \alpha \, a^2$. Since $\alpha \supseteq \gamma$, $c \, \gamma \, a$ implies $c \, \alpha \, a$, so that $bc \, \alpha \, ba \, \alpha \, a^2 = aa$. By ii), $b \, \alpha \, a$, but if $b \, \alpha \, a$ and $b \, \beta \, a$, then $b(\alpha \cap \beta)a$, and b is an element of $R(a, \alpha) \cap R(a, \beta)$, or $g = bc$ is an element of $R[a^2, (\alpha \cap \beta) \cup \gamma]$. It follows that

$$R[a^2, \alpha \cap (\beta \cup \gamma)] \subseteq R[a^2, (\alpha \cap \beta) \cup \gamma].$$

Lemma 1.4 implies that $\alpha \cap (\beta \cup \gamma) \subseteq (\alpha \cap \beta) \cup \gamma$.

2. Normal divisors. If H is a quasigroup contained in G, H is a normal divisor of G if and only if for some normal relation α, $H = R(a, \alpha)$ where a is any element of H. By Lemma 1.4, α is unique. We shall say that α is a normal divisor relation at a.

LEMMA 2.1. *If α is a normal relation on G and if $R(a, \alpha)R(a, \alpha) = R(a, \alpha)$, then α is a normal divisor relation at a.*

If $b \, \alpha \, a$, then $ba \, \alpha \, a$ by the closure of $R(a, \alpha)$. Thus, if $bx = c$ where b and c are elements of $R(a, \alpha)$, then $bx \, \alpha \, c \, \alpha \, a \, \alpha \, ba$, and by i) $x \, \alpha \, a$, or x lies in $R(a, \alpha)$. Similarly the solution of the equation $yb = c$ lies in $R(a, \alpha)$, and $R(a, \alpha)$ is a quasigroup.

LEMMA 2.2. *If α is a normal divisor relation at a, and if β is a normal relation such that $\beta \supseteq \alpha$, then β is a normal divisor relation at a.*

By Theorem 1.2

$$R(a, \beta)R(a, \beta) = R(a^2, \beta).$$

Since $R(a, \alpha)$ is a quasigroup, $a^2 \, \alpha \, a$, and since $\beta \supseteq \alpha$, $a^2 \, \beta \, a$. Then $R(a^2, \beta) = R(a, \beta)$, and by Lemma 2.1 $R(a, \beta)$ is a quasigroup.

If H and K are quasigroups contained in G, $H \cup K$ is the least quasigroup containing both H and K. Obviously if HK is a quasigroup containing H and K, then $HK = H \cup K$.

LEMMA 2.3. *If α and β are normal divisor relations at a, then*

$$R(a, \alpha \cup \beta) = R(a, \alpha)R(a, \beta) = R(a, \alpha) \cup R(a, \beta).$$

By Lemma 2.2 $R(a, \alpha \cup \beta)$ is a quasigroup, for $\alpha \cup \beta \supseteq \alpha$. Thus by Theorem 1.2

$$R(a, \alpha \cup \beta) = R(a^2, \alpha \cup \beta) = R(a, \alpha)R(a, \beta).$$

Clearly $R(a, \alpha)$ and $R(a, \beta)$ are contained in this set since $\alpha \cup \beta \supseteq \alpha$ and $\alpha \cup \beta \supseteq \beta$.

THEOREM 2.1. *If α and β are normal divisor relations at a, then $\alpha \cup \beta$ and $\alpha \cap \beta$ are normal divisor relations at a, and*

$$R(a, \alpha \cup \beta) = R(a, \alpha) \cup R(a, \beta),$$
$$R(a, \alpha \cap \beta) = R(a, \alpha) \cap R(a, \beta).$$

The theorem follows from Lemma 2.3, Theorem 1.2, and the fact that the intersection of two quasigroups is a quasigroup.

COROLLARY. *The set of normal divisor relations at a where a is an element of G constitutes a modular lattice L_a. The lattice L_a is isomorphic with the set of normal divisors of G which contain the element a.*

If β is a normal relation on G, and if K is a quasigroup contained in G, β is, by Lemma 1.3, a normal relation on K. The equivalence classes $R_K(g, \beta)$ constitute a quasigroup which we designate by K/β. If H is a quasigroup contained in K, and if $H \cap R_K(g, \beta)$ is not empty for every g in K, then $H/\beta \cong K/\beta$ under the correspondence of $H \cap R_K(g, \beta)$ to $R_K(g, \beta)$.

Now let α be a normal divisor relation on G at a, and let β be a normal relation on G. Then by Lemma 2.2 $\alpha \cup \beta$ is a normal divisor relation at a, and the quasigroup $R(a, \alpha)$ is contained in the quasigroup $R(a, \alpha \cup \beta)$. By Theorem 1.2 $R(a, \alpha \cup \beta) = R(a, \alpha)R(a, \beta)$. Let g be an element of $R(a, \alpha \cup \beta)$: Then $g = cd$ where $c \,\alpha\, a$, $d \,\beta\, a$. Then by iii) $ca \,\beta\, cd = g$, and if we designate $R(a, \alpha \cup \beta)$ by K, $R_K(g, \beta)$ contains the element ca which is an element of $R(a, \alpha)$. Thus $R_K(g, \beta) \cap R(a, \alpha)$ is never vacuous, and, as in the preceding paragraph,

$$R(a, \alpha \cup \beta)/\beta \cong R(a, \alpha)/\beta.$$

But $R(a, \alpha)/\beta = R(a, \alpha)/\alpha \cap \beta$ since the partition of β on $R(a, \alpha)$ is the same as that of $\alpha \cap \beta$.

THEOREM 2.2. *If α and β are normal relations on G such that $R(a, \alpha)$ is a quasigroup for some element a of G, then $R(a, \alpha \cup \beta)$ is a quasigroup, and*

$$R(a, \alpha \cup \beta)/\beta \cong R(a, \alpha)/\alpha \cap \beta.$$ [6]

[6] Cf. Garrison [10], Theorem 3.2.

If in Theorem 2.2 β is also a normal divisor relation at a, we have the familiar isomorphism theorem which yields through the Ore Theorem on chains in partially ordered sets [7] the Jordan-Hölder Theorem for composition series in G.[8]

3. Quasigroups containing idempotent elements. The element e of G is called idempotent if $e^2 = e$.

LEMMA 3.1. *If e is an idempotent element of G, every normal relation on G is a normal divisor relation at e.*[9]

We may employ Theorem 1.2 and Lemma 2.1 to show that $R(e, \alpha)$ is a quasigroup when α is a normal relation; for

$$R(e, \alpha)R(e, \alpha) = R(e^2, \alpha) = R(e, \alpha).$$

THEOREM 3.1. *If G contains an idempotent element e, the lattice L of normal relations on G is isomorphic with the lattice of normal divisors of G containing e.*

COROLLARY. *If G contains the idempotent elements e_1 and e_2, then $L_{e_1} \cong L_{e_2}$.*

COROLLARY. *The normal divisors of any loop form a modular lattice.*[10] *Further, if a loop G is homomorphic to a quasigroup $\bar{G}$, there will exist a normal divisor H of G such that $G/H \cong \bar{G}$.*

The direct product of quasigroups $A_1, A_2, \cdots, A_n$ is a quasigroup $G = A_1 \times A_2 \times \cdots \times A_n$ whose elements are the n-tuples $(a_1, a_2, \cdots, a_n)$, a_i in A_i, $i = 1, 2, \cdots, n$.[11] The ordered product of two elements is the n-tuple of ordered products of corresponding components. If A_i contains the idempotent elements e_i, $i = 1, 2, \cdots, n$, then G contains the idempotent element $e = (e_1, e_2, \cdots, e_n)$. Conversely, if e is idempotent in G and if $e = (e_1, e_2, \cdots, e_n)$, then e_i is idempotent in A_i.

Obviously the quasigroup A_i may be identified with the quasigroup

$$e_1 \times e_2 \times \cdots \times e_{i-1} \times A_i \times e_{i+1} \times \cdots \times e_n$$

[7] See Ore [14].

[8] Cf. Garrison [10], Theorem 3.3.

[9] Cf. Garrison [10], Lemma 3.2.

[10] Smiley has established this theorem in [15].

[11] See Bruck [6], p. 48 for this definition.

which is a sub-quasigroup of G, and the elements $e_1, e_2, \cdots, e_n$ may be identified with the element e. Thus G is the direct product of sub-quasigroups whose intersection in pairs is the element e. Under these conditions we shall say that the decomposition is direct over e. It follows trivially that the A_i are normal in G.

THEOREM 3.2. *Let G contain the idempotent element e, and let α and β be normal relations on G. If $G = R(e,\alpha)R(e,\beta)$ and $R(e,\alpha) \cap R(e,\beta) = e$, then $G = R(e,\alpha) \times R(e,\beta)$.*[12]

Let a lie in $R(e,\alpha)$ and b in $R(e,\beta)$. Define $\bar{G} = R(e,\alpha) \times R(e,\beta)$ as above, and let θ be the mapping defined by $(ab)^\theta = (ae, eb)$. This is a bi-unique mapping of G on $\bar{G}$. Consider that if a_1b_1 and a_2b_2 are two elements of G, then $(a_1b_1)(a_2b_2) = ab$ for some a and b. Since $a_1b_1 \beta a_1e$, $a_2b_2 \beta a_2e$, and $ab \beta ae$, it follows that $(a_1e)(a_2e) \beta ae$. All of these elements lie in $R(e,\alpha)$, so that $(a_1e)(a_2e) \alpha ae$, and thus $(a_1e)(a_2e) (\alpha \cap \beta) ae$. Since $R(e,\alpha) \cap R(e,\beta) = e = R(e,\alpha \cap \beta)$, $\alpha \cap \beta$ is equality by Lemma 1.4, and $(a_1e)(a_2e) = ae$. Similarly $(eb_1)(eb_2) = eb$. Then

$$\begin{aligned}(a_1b_1)^\theta(a_2b_2)^\theta &= (a_1e, eb_1)(a_2e, eb_2)\\ &= (ae, eb)\\ &= (ab)^\theta\\ &= [(a_1b_1)(a_2b_2)]^\theta,\end{aligned}$$

and θ is an isomorphism.

BIBLIOGRAPHY

1. A. A. Albert, "Quasigroups. I," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 507-519.
2. ———, "Quasigroups. II," *Transactions of the American Mathematical Society*, vol. 55 (1944), pp. 401-419.
3. R. Baer, "The homomorphism theorems for loops," *American Journal of Mathematics*, vol. 67 (1945), pp. 450-460.
4. G. Birkhoff, *Lattice theory*, American Mathematical Society Colloquium Publications, vol. 25, 1940.

---

[12] This theorem has been established by Albert for loops in [2], p. 409.

5. R. H. Bruck, "Some results in the theory of linear non-associative algebras," *Transactions of the American Mathematical Society*, vol. 56 (1944), pp. 141-199.
6. ———, "Some results in the theory of quasigroups," *Transactions of the American Mathematical Society*, vol. 55 (1944), pp. 19-52.
7. ———, "Simple quasigroups," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 769-781.
8. ———, "Contributions to the theory of loops," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 245-354.
9. G. N. Garrison, "Quasigroups," *Annals of Mathematics*, vol. 41 (1940), pp. 474-487.
10. ———, "Note on invariant complexes of a quasigroup," *Annals of Mathematics*, vol. 47 (1946), pp. 50-55.
11. B. A. Hausmann and O. Ore, "Theory of quasigroups," *American Journal of Mathematics*, vol. 59 (1937), pp. 983-1004.
12. D. C. Murdoch, "Note on normality in quasigroups," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 134-138.
13. O. Ore, "Theory of equivalence relations," *Duke Mathematical Journal*, vol. 9 (1942), p. 573.
14. ———, "Chains in partially ordered sets," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 558-566.
15. M. F. Smiley, "An application of lattice theory to quasigroups," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 782-786.

THE TORSIONAL RIGIDITY AND VARIATIONAL METHODS.*

By J. B. Diaz and A. Weinstein.

The present paper consists of two independent remarks concerning the rigidity of beams in torsion. The first part contains a method for the computation of the torsional rigidity in the case of multiply connected domains. The second part deals with an application of the Rayleigh-Ritz method to the torsion of beams of simply connected cross section.

1. An extremal property of the torsional rigidity. Let $\phi(x, y)$ be the warping function, and $\psi(x, y)$ its conjugate, in the problem of a beam, with simply or multiply connected cross section, in torsion. As is well known, ϕ is the solution of the following Neumann problem:

(1) $$\Delta\phi = 0, \text{ in the cross section } R,$$

(2) $$\frac{\partial \phi}{\partial n} = yn_x - xn_y, \text{ on the boundary } C \text{ of } R,$$

where n_x and n_y denote the components of the external normal n. For a multiply connected domain, C consists of a simple closed curve C_0 and a finite number of simple closed curves $C_1, \cdots, C_m$, lying in the interior of C_0.

The conjugate harmonic function ψ is the solution of the following Dirichlet problem:

(3) $$\Delta\psi = 0, \qquad \text{in } R,$$

(4) $$\psi = (1/2)(x^2 + y^2), \qquad \text{on } C_0,$$

(5) $$\psi = (1/2)(x^2 + y^2) + k_i, \qquad \text{on } C_i,$$

where the constants k_i are not explicitly given a priori. This introduces a serious complication in the torsion of multiply connected domains.

The torsional rigidity, or stiffness, S, can be expressed in several ways, either in terms of ϕ or ψ, in the form of an integral taken over R. This is one of the rare cases, in statics, where the actual numerical value of an integral of an unknown function is of major interest. Usually, S is given by the formula [1]

* Received March 1, 1947; Presented to the American Mathematical Society, December 27, 1946.

[1] Some authors define μS as the torsional rigidity, μ being Lamé's constant. Another expression for S, due to Prandtl, will be considered in 4.

$$(6) \qquad S = \int\int_R [x^2 + y^2 + x\phi_y - y\phi_x]\, dxdy.$$

Using the Cauchy-Riemann equations

$$(7) \qquad \phi_x = \psi_y, \qquad \phi_y = -\psi_x,$$

an analogous expression for S, in terms of ψ, is obtained.

However, for our purposes, it is convenient to employ another formula for S; which, although closely connected with (6), has not, to our knowledge, been used explicitly in the theory of torsion. By Green's theorem, using (2)

$$\int\int_R [x\phi_y - y\phi_x]\, dxdy = \int\int_R \left[\frac{\partial(x\phi)}{\partial y} - \frac{\partial(y\phi)}{\partial x}\right] dxdy$$

$$= \int_C \phi[xn_y - yn_x]\, ds = -\int_C \phi \frac{\partial\phi}{\partial n}\, ds = -D(\phi),$$

where

$$D(\phi) = \int\int_R [\phi^2_x + \phi^2_y]\, dxdy,$$

is the Dirichlet integral. Denoting by

$$(8) \qquad P = \int\int_R (x^2 + y^2)\, dxdy,$$

the polar moment of inertia of the section R, we have, in place of (6), the *fundamental formula*

$$(9) \qquad S = P - D(\phi).$$

Observing that, by (7), $D(\phi) = D(\psi)$, we have also

$$(10) \qquad S = P - D(\psi).$$

Since $D(\phi) \geqq 0$, we have from (6) the following extremal property of the torsional rigidity, considered as a function of the section.

For any (simply or multiply connected) section

$$(11) \qquad S \leqq P.$$

The equality sign holds only when $D(\phi) = 0$, which implies that the warping function is a constant. As is well known, this can only happen if R is a circular region or a circular ring, bounded by two concentric circles.

We note the curious fact that for domains bounded by more than two curves there is no section for which S equals the polar moment of inertia of the section.

2. Upper and lower bounds for the Dirichlet integral in Neumann's problem. Besides yielding the inequality (11), formulas (9) and (10) are adapted for obtaining estimates of the torsional rigidity. In particular, (9) is suited for multiply connected domains, a case which presented great difficulties in the past. In fact, upper and lower bounds for $D(\phi)$ yield lower and upper bounds, respectively, for S.

An upper bound for $D(\phi)$ in Neumann's problem is given by Kelvin's theorem.[2] Let p and q be the components of a divergence-free vector; i. e.,

$$p_x + q_y = 0, \qquad \text{in } R,$$

satisfying the condition

$$pn_x + qn_y = \frac{\partial \phi}{\partial n}, \qquad \text{on } C,$$

(where, in our case, $\frac{\partial \phi}{\partial n}$ is given by (2)). Then,

$$D(\phi) \leqq \iint_R [p^2 + q^2]\,dxdy. \tag{12}$$

Let us note that an estimate of this kind can always be improved by taking for (p, q) a vector field containing one or several parameters (Rayleigh-Ritz method).

The inequality (12), originally employed in three dimensions by Kelvin, can be greatly simplified in the two-dimensional case by the introduction of a "stream function" $K(x, y)$ defined by

$$K_x = -q, \qquad K_y = p. \tag{13}$$

The function K, which is arbitrary in the interior of the cross section, satisfies the following boundary condition

$$K_y n_x - K_x n_y = \frac{\partial \phi}{\partial n} = \frac{d}{ds}\,[\tfrac{1}{2}(x^2 + y^2)].$$

The left hand member of the last equation is obviously $\frac{dK}{ds}$, so that we have, on the boundary of R:

$$\frac{dK}{ds} = \frac{d}{ds}\,[\tfrac{1}{2}(x^2 + y^2)]. \tag{14}$$

---

[2] See, for example, H. Lamb, *Hydrodynamics*, Sixth Edition, New York, 1945, p. 47. The theorem obviously holds as well for multiply connected domains.

Hence K may be any continuous function, with continuous derivatives, satisfying the boundary conditions

$$K(x,y) = (1/2)(x^2 + y^2) + c_i, \qquad \text{on } C_i, \tag{15}$$

$i = 0, \cdots, m$, where the constants c_i are perfectly arbitrary. It is interesting to note that the manifold of the functions K contains the function ψ; which, by (3), (4), and (5), can be obtained by taking K to be harmonic and choosing the constants c_i in (15) in such a way that the conjugate function ϕ is single valued. Of course, as remarked previously, the precise values of the desired constants are not known, except in trivial cases.

Using K, (12) may be replaced by [3]

$$D(\phi) \leqq D(K). \tag{16}$$

Generally speaking, for any fixed choice of the constants c_i in (15) we are still at liberty to choose K in the interior, and the smallest upper bound for $D(\phi)$ in (16) is obviously obtained by taking for K the harmonic function K_h corresponding to the given boundary values. Consequently,

$$D(\phi) \leqq D(K_h) \leqq D(K). \tag{17}$$

This remark will be useful in applications.

We turn now to the determination of a lower bound for $D(\phi)$ in Neumann's problem, which will be obtained by a simple application of Schwarz's inequality and a transformation involving Green's formula. This procedure is equivalent to what is sometimes called Trefftz' method,[4] but is considerably simpler, and also perfectly general, as will be shown elsewhere.

Our starting point is Schwarz's inequality

$$[D(f,\phi)]^2 \leqq D(f)D(\phi),$$

which we use in the equivalent form

$$\left[\frac{D(f,\phi)}{D(f)}\right]^2 D(f) \leqq D(\phi),$$

so that

$$D(\lambda f) \leqq D(\phi), \tag{18}$$

where

[3] Since $D(\phi) = D(\psi)$, we have also $D(\psi) \leqq D(K)$. This last inequality could have been obtained by minimizing $D(K)$, where K satisfies (15), and the c_i's are variable parameters.

[4] E. Trefftz, "Ein Gegenstuck zum Ritzchen Verfahren," *Verhandlungen, Kongress für technische Mechanik*, Zürich, 1927, p. 131. Some time ago, Professor J. L. Synge kindly communicated to us another computation of estimates for $D(\phi)$ in Neumann's problem.

$$\lambda = \frac{D(f, \phi)}{D(f)}. \tag{19}$$

Let us observe that the quantity λ is known, as we have, by Green's formula,

$$D(f, \phi) = -\iint_R f \cdot \Delta\phi dxdy + \int_C f \frac{\partial \phi}{\partial n} ds = \int_C f \frac{\partial \phi}{\partial n} ds. \tag{20}$$

Thus, if

$$\int_C f \frac{\partial \phi}{\partial n} ds \neq 0, \tag{21}$$

the quantity $D(\lambda f)$ in (18) is a positive lower bound for $D(\phi)$. Since $\frac{\partial \phi}{\partial n}$ is given on C, there are infinitely many functions satisfying condition (21). Again the result can be improved by introducing arbitrary parameters into f.

In view of the bounds (16) and (18) for $D(\phi)$, equation (9) yields finally

$$P - D(K) \leqq S \leqq P - \frac{\left(\int_C f \frac{\partial \phi}{\partial n} ds\right)^2}{D(f)}, \tag{22}$$

where $\frac{\partial \phi}{\partial n}$ is given by (2), so that

$$\int_C f \frac{\partial \phi}{\partial n} ds = \iint_R [yf_x - xf_y] dxdy.$$

Therefore, we have finally

$$P - D(K) \leqq S \leqq P - \frac{\left(\iint_R [yf_x - xf_y] dxdy\right)^2}{D(f)}, \tag{23}$$

where the function K satisfies (15), and the function f satisfies (21). This double inequality is the source of many interesting general statements. Taking for f the imaginary part of $(x + iy)^n$, we obtain

$$S \leqq P - \frac{\left(\iint_R \operatorname{Re}(x + iy)^n dxdy\right)^2}{\iint_R (x^2 + y^2)^{n-1} dxdy},$$

where Re denotes the real part. In particular, for $n = 2$ it follows easily that

$$S \leqq \frac{4AB}{A + B},$$

where A and B are the moments of inertia of R with respect to the x and y axes, and $P = A + B$.[5]

---

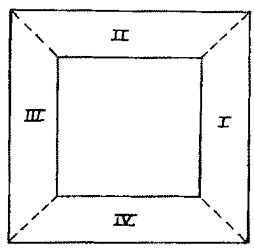
[5] Cf. a brief note by E. Nicolai, *Zeitschrift für angewandte Mathematik*, vol. 4 (1924), p. 181.

3. Application. As a simple application we derive an inequality for the torsional rigidity of the quadratic frame bounded by the lines $x = \pm a$, $y = \pm b$, $0 < a < b$. Here the boundary C consists of the outer square C_0 and the inner square C_1.[6]

According to the general theory, we may take for K any function assuming on C_0 the same boundary values as $(1/2)(x^2 + y^2) - (1/2)b^2$ and on C_1 the same values as $(1/2)(x^2 + y^2) - (1/2)a^2$. Let K_h be the harmonic function which takes these boundary values. By (17), we have

$$D(\phi) \leqq D(K_h).$$

Exploiting the symmetry of the domain, we may subdivide it into four quadrangles I, II, III, IV. Let K^* be defined as follows:



$$K^*(x, y) = \begin{cases} (1/2)y^2, & \text{in I and III,} \\ (1/2)x^2, & \text{in II and IV.} \end{cases}$$

Clearly K^* has the boundary values prescribed for K. The function K^* is continuous in the cross section, but its first derivatives have jumps along the four diagonals indicated in the figure. By the general theory of Dirichlet's problem, we have

$$D(K_h) \leqq D(K^*), \tag{24}$$

hence

$$D(\phi) \leqq D(K^*). \tag{25}$$

For the domain under consideration

$$P = (8/3)(b^4 - a^4),$$

and a glance at the figure shows that

$$D(K^*) = 4 \int\int_I y^2 \, dx dy = (2/3)(b^4 - a^4).$$

---

[6] Cf. also a brief note by C. Weber, *Zeitschrift für angewandte Mathematik*, vol. 11 (1931), p. 244.

Combining the last two equations with (11) and (23), we have finally

$$2(b^4-a^4) \leqq S \leqq (8/3)(b^4-a^4). \tag{26}$$

4. The torsional rigidity and the minimum of the potential energy.[7] It is well known that the torsional rigidity is related to the minimum of the energy integral. Let us consider a beam in torsion whose simply connected cross section R is bounded by a closed curve C. Let $\Psi = \psi - (1/2)(x^2+y^2)$ be Prandtl's stress function, satisfying the differential equation

$$\Delta\Psi = -2, \qquad \text{in } R, \tag{27}$$

and the boundary condition

$$\Psi = 0, \qquad \text{on } C. \tag{28}$$

The torsional rigidity, S, is then given by the formulas

$$S = D(\Psi), \qquad S = 2\iint_R \Psi\, dxdy. \tag{29}$$

Among all functions u satisfying the boundary condition

$$u = 0, \qquad \text{on } C, \tag{30}$$

the function Ψ yields the minimum value for the integral

$$I(u) = \iint_R [u_x{}^2 + u_y{}^2 - 4u]\,dxdy. \tag{31}$$

Since, by Green's formula

$$I(\Psi) = -\iint_R \Psi(\Delta\Psi)\,dxdy - 4\iint_R \Psi\,dxdy$$
$$= -2\iint_R \Psi\,dxdy,$$

it follows from (29) and (31) that the torsional rigidity and the minimum value of I are connected by the relations

$$D(\Psi) = -I(\Psi), \qquad 2\iint_R \Psi\,dxdy = -I(\Psi). \tag{32}$$

The derivation of formulas (32) depends essentially on the fact that Ψ satisfies not only the boundary condition (28): $\Psi = 0$ on C, but also the differential equation (27):

$$\Delta\Psi = -2, \qquad \text{in } R.$$

[7] This section is independent of the preceding considerations.

It in no way follows that (32) will continue to hold if Ψ is replaced by a function which is only required to satisfy the boundary condition (28), as is demanded in the application of the Rayleigh-Ritz method. The main purpose of this section is to show that the following equalities (analogous to (32))

$$(33)\qquad D(\Psi^*_N) = -I(\Psi^*_N), \qquad 2\int\int_R \Psi^*_N \, dxdy = -I(\Psi^*_N),$$

hold,[8] provided that Ψ^*_N is the *best approximation* to Ψ in the sense of the Rayleigh-Ritz method, obtained by using N fixed functions $f_i(x, y)$ satisfying the boundary conditions

$$f_i = 0, \qquad \text{on } C,$$

for $i = 1, \cdots, N$.

Let

$$(34)\qquad \Psi_N(c_1, \cdots, c_N) = \sum_{i=1}^{N} c_i f_i(x, y),$$

in R. Then, for a fixed choice of the f_i's the integral I is a function of the parameters $c_1, \cdots, c_N$, and will be denoted by $I\ (c_1, \cdots, c_N)$. For all values of $c_1, \cdots, c_N$

$$(35)\qquad I(c_1, \cdots, c_N) \geqq I(\Psi).$$

The *best upper bound* (obtainable from the f_i's started with) for $I(\Psi)$, is given by the minimum value of $I(c_1, \cdots, c_N)$. Denote by $c^*_1, \cdots, c^*_N$ the values of the c_i's for which the minimum value is attained. The c^*_i's are given explicitly by the solution of the N linear equations

$$(36)\qquad \frac{\partial I}{\partial c_h} = 0, \qquad h = 1, \cdots, N,$$

and, in accordance with our previous notation

$$(37)\qquad \Psi^*_N = \sum_{i=1}^{N} c^*_i f_i(x, y).$$

For any values of the c_i's

$$(38)\qquad I(c_1, \cdots, c_N) = \sum_{i,k=1}^{N} a_{ik} c_i c_k + \sum_{j=1}^{N} b_j c_j,$$

where

$$(39)\qquad a_{ik} = \int\int_R \left[\frac{\partial f_i}{\partial x}\frac{\partial f_k}{\partial x} + \frac{\partial f_i}{\partial y}\frac{\partial f_k}{\partial y}\right] dxdy,$$

$(a_{ik} = a_{ki})$, and

$$(40)\qquad b_j = -4\int\int_R f_j \, dxdy.$$

---

[8] In other words, we have $D(\Psi^*_N) = 2\int\int_R \Psi^*_N \, dxdy$, which is analogous to the equation $D(\Psi) = 2\int\int_R \Psi \, dxdy$ for Ψ. Clearly the relation $D(u) = 2\int\int_R u \, dxdy$ is not satisfied by every function u which vanishes on the boundary of R.

But, from Euler's theorem for homogeneous functions,

$$2\sum_{i,k=1}^{N} a_{ik}c_ic_k = \sum_{h=1}^{N} c_h \frac{\partial[\sum_{i,k=1}^{N} a_{ik}c_ic_k]}{\partial c_h}. \tag{41}$$

Employing (38), (41) gives

$$2[I - \sum_{j=1}^{N} b_jc_j] = \sum_{h=1}^{N} c_h \frac{\partial I}{\partial c_h} - \sum_{h=1}^{N} c_hb_h,$$

or

$$I(c_1, \cdots, c_N) = \tfrac{1}{2}\sum_{h=1}^{N} c_h \frac{\partial I}{\partial c_h} + \tfrac{1}{2}\sum_{j=1}^{N} b_jc_j. \tag{42}$$

For $c^*_i, \cdots, c^*_N$, in view of (36), equation (42) yields

$$I(c^*_1, \cdots, c^*_N) = \tfrac{1}{2}\sum_{j=1}^{N} b_jc^*_j; \tag{43}$$

which, together with (37) and (40), furnishes the second equation of (33). Recalling the definition of I, equation (31), we have

$$2\iint_R \Psi^*_N \, dxdy = D(\Psi^*_N),$$

and the first equation of (33) follows.

For multiply connected domains, the torsional rigidity is given by the formulas (cf. (29))

$$S = D(\Psi), \qquad S = 2\iint_R \Psi \, dxdy + 2\sum_{i=1}^{m} k_iA_i, \tag{44}$$

the k_i being the constants appearing in (5), A_i being the area enclosed by the curve C_i, and $\Psi = \psi - \frac{1}{2}(x^2 + y^2)$. The function Ψ minimizes the energy integral $I(u)$ over the class of all functions u which satisfy the boundary conditions for Ψ, i. e., which vanish on C_0 and assume the constant values k_i on C_i. Furthermore, we have

$$\begin{aligned} D(\Psi) - 4\sum_{i=1}^{m} k_iA_i &= -I(\Psi), \\ 2\iint_R \Psi \, dxdy - 2\sum_{i=1}^{m} k_iA_i &= -I(\Psi), \end{aligned} \tag{45}$$

which replaces (32).

One would be tempted to obtain Ψ by minimizing $I(u)$ over the class of all functions u which vanish on C_0 and assume arbitrary constant values c_i

on C_i (the c_i being variable parameters in the minimum problem). It is easy to see that in this case the minimizing function U will satisfy

$$\Delta U = -2, \qquad \text{in } R,$$

$$U = \gamma_i, \qquad \text{on } C_i, \tag{46}$$

$$\int_{C_i} \frac{\partial U}{\partial n}\, ds = 0, \qquad \text{on } C_i,$$

$i = 1, \cdots, m$, where the γ_i's are certain constants. It follows at once that for a multiply connected domain R, U cannot coincide with Ψ, since

$$\int_{C_i} \frac{\partial \Psi}{\partial n}\, ds = 2A_i, \tag{47}$$

$i = 1, \cdots, m$. Hence

$$I(U) < I(\Psi). \tag{48}$$

For this reason, the Rayleigh-Ritz method cannot be applied to the integral $I(u)$, using functions u which vanish on C_0 and are constant on the C_i, inasmuch as one would then obtain upper bounds for $I(U)$ and not $I(\Psi)$. This fact shows clearly the advantage of the method developed in the first part of this paper.[9]

Brown University.
Carnegie Institute of Technology.

[9] However it should be noted that Courant (*Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 1-23, especially pp. 6, 7 and 20, 21) obtains lower bounds for S by minimizing $I(u) - 4\Sigma c_i A_i$ over the class of functions u vanishing on C_0 and assuming arbitrary constant values c_i on C_i.

According to a communication of Professor T. J. Higgins of May 29, 1947 the only place where a formula equivalent to (10) seems to have been mentioned (for simply connected domains) is a Thesis by M. Stone, "Electrical Analogy in the Study of Torsion," Abstract of Thesis, Graduate School, University of Pittsburgh, vol. 9 (1933), pp. 265-274. For further literature on torsional rigidity, see T. J. Higgins, "The Approximate Mathematical Methods of Applied Physics as Exemplified by Application to Saint-Venant's Torsion Problem," *Journal of Applied Physics*, vol. 14 (1943), pp. 469-480.

SEMIGROUPS HAVING ZEROID ELEMENTS.*

By A. H. CLIFFORD and D. D. MILLER.

By a *semigroup* we shall mean a system consisting of a set S in which a single-valued product ab is defined for every pair a, b of S, and for which the associative law holds:

$$(ab)c = a(bc).$$

An element u of S will be called a *zeroid element* of S if, for each element a of S, there exist x and y in S such that $ax = ya = u$. According to Huntington's set of group axioms,[1] S is a group if and only if every element of S is a zeroid. If S has a zero element, e. g. if S is the multiplicative semigroup of a ring, then zero is the only zeroid element of S.

In the present paper we show that the set U of zeroid elements of any semigroup S is either vacuous or else is a subgroup of S. U is a two-sided ideal contained in every left, right or two-sided ideal of S. It is therefore the "Kerngruppe" of S in the sense of Suschkewitsch.[2] The identity element z of U commutes with every element of S, and the mapping $a \rightarrow za$ $(= az)$ is a homomorphism of S onto U. We define the *core* J of S to be the set of elements mapped into z. J is a subsemigroup of S containing z as zero element. If, on the other hand, we start with a group U and a semigroup J with zero, we can construct at least one semigroup S (e. g. the direct product of U and J) such that the group of zeroid elements of S is isomorphic with U.

A few properties of subgroups and subsemigroups of a semigroup S having zeroid elements are discussed, and it is noted that any semigroup homomorphic with a subsemigroup of a group can be embedded in such a

---

* Received February 28, 1947.

[1] E. V. Huntington, "Simplified definition of a group," *Bulletin of the American Mathematical Society*, vol. 8 (1901-02), pp. 296-300. As a matter of fact, Huntington proves closure as a theorem from associativity and the existence of solutions x and y of $ax = b$ and $ya = b$. We do not really need this refinement in Lemma 2 since closure is immediate.

[2] A. Suschkewitsch, "Über die endlichen Gruppen ohne das Gesetz der eindeutige Umkehrbarkeit," *Mathematische Annalen*, vol. 99 (1928), pp. 30-50. The description of S as the intersection of all the two-sided ideals of S is due to D. Rees, "On semi-groups," *Proceedings of the Cambridge Philosophical Society*, vol. 36 (1940), pp. 387-400, esp. p. 392. The fact that U is a group is a consequence of Suschkewitsch's work, but he deals only with finite S. His results will be extended to infinite S in a later paper.

semigroup S. Homomorphisms of a semigroup onto a group, or onto a semigroup with cancellation, have been studied by Dubreil.[3]

1. Right and left zeroids; minimal right and left ideals. An element r of a semigroup S will be called a *right zeroid element* of S if, for each element a of S, there exists at least one element x of S such that $ax = r$. By a *right ideal* in S we shall mean a non-vacuous subset R of S such that $RS \subseteq R$. The set R of all right zeroid elements of S is a right ideal, provided it is not vacuous. For if $r \varepsilon R$ and $s \varepsilon S$, and we solve $ax = r$ for x, then the equation $ax_1 = rs$ has the solution $x_1 = xs$. R is contained in every right ideal A of S; for if $r \varepsilon R$ and $a \varepsilon A$, and we solve $ax = r$, it is immediate that $r \varepsilon A$ from the definition of right ideal.

We shall call a right ideal *universally minimal* in S if it is contained in every right ideal of S, and *locally minimal* if it contains no proper right subideal. As an example of a semigroup containing locally minimal right ideals but no universally minimal right ideal, consider the set S_4 of the four elements a_{11}, a_{12}, a_{21}, a_{22}, with multiplication defined as follows:

$$a_{ij}a_{kl} = a_{il}.$$

The set $\{a_{11}, a_{12}\}$ is a locally minimal right ideal in S_4, and the set $\{a_{21}, a_{22}\}$ is another. Since the intersection of these two sets is vacuous, S_4 contains no universally minimal right ideal.

Suppose now that a semigroup S contains a universally minimal right ideal R. If a is any element of S, the set aS of all ax with x in S is a right ideal (the "principal right ideal generated by a"). By hypothesis on R, $aS \supseteq R$. Hence if $r \varepsilon R$ there is an x such that $ax = r$, whence r is a right zeroid element of S. Consequently we see that a semigroup S contains a right zeroid element if and only if it contains a universally minimal right ideal R; and then R consists of all the right zeroid elements of S.

Similarly, we define a *left zeroid element* of S to be an element l of S such that, for each element a of S, there exists at least one $y \varepsilon S$ such that $ya = l$. S contains a left zeroid element if and only if it contains a universally minimal left ideal L; and then L consists of all the left zeroid elements of S.

The set U of all (two-sided) zeroid elements of S is clearly the inter-

[3] Paul Dubreil, "Contribution à la théorie des demi-groupes," *Mémoires de l'Académie des Sciences de l'Institut de France* (2), vol. 63 (1941), pp. 1-52, especially pp. 27-28.

section $R \cap L$. If R and L are non-vacuous, then U is non-vacuous, since it contains the set RL of all products rl $(r \varepsilon R,\ l \varepsilon L)$.

The example S_4 given above has no right zeroids, since it has no universally minimal right ideal. Similarly, it has no left zeroids, since it contains the locally minimal left ideals $\{a_{11}, a_{12}\}$ and $\{a_{12}, a_{22}\}$ with vacuous intersection. Thus S_4 contains no zeroid elements of either kind. A semigroup may contain one kind but not the other; for example, one of the minimal right ideals of S_4 is a semigroup every element of which is a right zeroid, but containing no left zeroid. *In what follows we make the basic assumption that S contains at least one left and one right zeroid, and hence at least one (two-sided) zeroid.*

2. Proof that U is a group, and that every one-sided zeroid is two-sided.

We owe to R. H. Bruck the idea of first proving that the set $V = RL$ is a group, which greatly shortens our original proof.

LEMMA 1R. *If $a \varepsilon S$ and $t \varepsilon V$, the equation $ax = t$ has a solution x in V.*

Proof. aR is a right ideal in S. Since R is universally minimal, $aR \supseteq R$, and hence

$$aV = aRL \supseteq RL = V.$$

LEMMA 1L. *If $a \varepsilon S$ and $t \varepsilon V$, the equation $ya = t$ has a solution y in V.*

Proof. La is a left ideal, whence $La \supseteq L$ and $RLa \supseteq RL$.

LEMMA 2. *V is a group.*

Proof. $R \cdot LR \subseteq R$, whence $RL \cdot RL \subseteq RL$; i. e., $V = RL$ is a subsemigroup of S. From Lemmas 1R and 1L (true in particular for $a \varepsilon V$) we conclude that the Huntington axioms [1] are satisfied.

LEMMA 3. *Let z be the identity element of the group $V = RL$. Then $R = zS$, $L = Sz$, $U = zSz$.*

Proof. zS is a right ideal and hence contains R. But $z \varepsilon RL \subseteq R$, so that $zS \subseteq R$, whence equality follows. By left-right duality, $Sz = L$. Since z is idempotent, what we have shown is that R consists of all $x \varepsilon S$ such that $zx = x$, and L of all $y \varepsilon S$ such that $yz = y$. $U = R \cap L$ therefore consists of all x for which z is a two-sided identity element, or equivalently $U = zSz$.

LEMMA 4. $U = R = L = V$.

Proof. Let $a \varepsilon S$. By Lemma 1R, there exists $v \varepsilon V$ such that $av = z$ (z the identity element of V). If v^{-1} is the inverse of v in the group V we have

$$az = avv^{-1} = zv^{-1} = v^{-1} \varepsilon V.$$

Hence $Sz \subseteq V$, whence $L \subseteq V$ by Lemma 3. But $V = RL \subseteq L$, whence equality follows. Dually, $R = V$ and hence $U = R \cap L = V \cap V = V$.

THEOREM 1. *The set U of (two-sided) zeroid elements of a semigroup S is either vacuous or else a subgroup of S.*

Proof. Immediate from Lemmas 2 and 4.

THEOREM 2. *If a semigroup S contains at least one two-sided zeroid element, then every left zeroid is also a right zeroid and vice-versa; in other words, $R = L = U$.*

Proof. Immediate from Lemma 4.

Expressed in a slightly different way, Theorem 2 states that if a semigroup has universally minimal left and right ideals, then it has a minimal two-sided ideal, and all three coincide.

3. The homomorphism ζ of S onto U; the core and frame of S.

THEOREM 3. *Let S be a semigroup containing at least one two-sided zeroid element, and let z be the identity element of the group U of two-sided zeroid elements of S (Theorem 1). Then*

(1) *z commutes with every element of S;*

(2) *the mapping $a \rightarrow za$ $(= az)$ is a homomorphism ζ of S onto U;*

(3) *the set J of elements of S mapped into z by ζ is a subsemigroup of S containing z; J consists of all elements of S for which z is a zero element;*

(4) *the common part of U and J consists of z alone.*

Proof. (1) Let $a \varepsilon S$. By Theorem 2, $za \varepsilon U$ and $az \varepsilon U$. Since z is the identity element of U,

$$za = za \cdot z = z \cdot az = az.$$

(2) Since $za \,\varepsilon\, U$, ζ maps S into U. U is covered by ζ; in fact, U consists of all elements of S invariant under ζ. ζ is a homomorphism since

$$az \cdot bz = abz^2 = ab \cdot z.$$

(3) If $k_1 \,\varepsilon\, J$ and $k_2 \,\varepsilon\, J$, so that $zk_1 = z$ and $zk_2 = z$, then $zk_1k_2 = zk_2 = z$; therefore J is closed under multiplication, and hence is a subsemigroup of S. Since $zz = z$, $z \,\varepsilon\, J$. To say that z is a zero element for an element a of S means that $az = za = z$, i.e. $a \,\varepsilon\, J$.

(4) If $a \,\varepsilon\, U$ then $az = za = a$. If $a \,\varepsilon\, J$ then $az = za = z$. Hence if $a \,\varepsilon\, U \cap J$ then $a = z$.

We shall call J the *core* of S.

Theorem 3 gives us the following picture of the gross structure of S. To each element u of the group U corresponds a "congruence class" $J(u) \subseteq S$, consisting of those elements $x \,\varepsilon\, S$ such that $zx = xz = u$. Each element of S belongs to one and only one class $J(u)$. The class $J(u)$ contains one and only one element of U, namely u. The product of any element of $J(u_1)$ with any element of $J(u_2)$ is an element of $J(u_1u_2)$:

$$J(u_1) \cdot J(u_2) \subseteq J(u_1u_2).$$

If, moreover, either factor is in U then the product is in U; thus if $a_1 \,\varepsilon\, J(u_1)$ we have

$$a_1u_2 = u_1u_2 \text{ and } u_2a_1 = u_2u_1.$$

In this notation, the core of S is the class $J(z)$. The product of an element of $J(z)$ with an element of $J(u)$ in either order is an element of $J(u)$. Consequently, because of the associative law, the mappings

$$x \to kx \text{ and } x \to xk,$$

with $x \,\varepsilon\, J(u)$ and $k \,\varepsilon\, J(z)$, afford left and right representations of the core $J(z)$ by mappings of the set $J(u)$ into itself. Each left mapping commutes with each right mapping.

If $u \,\varepsilon\, U$ and $k \,\varepsilon\, J$ we have $uk = ku = u$; i.e. every element of U is a zero element for the elements of J. Consequently the class sum of U and J is a subsemigroup of S, which we shall call the *frame* of S.

4. Semigroups with a given frame. As remarked in the introduction, there is at least one semigroup having a preassigned group U for its group of zeroids, and for its core a preassigned semigroup J with zero, namely the direct product of U and J. In this section we shall give a simple construction

showing that there are an infinite number of such semigroups. We do not attempt the formidable task of finding them all.

The most economical construction is the following. Identify the identity element of U with the zero element of J, and define $uk = ku = u$ for every $u \varepsilon U$, $k \varepsilon J$. The frame of the resulting semigroup S_0 is S_0 itself. If S is any semigroup whose group of zeroids is isomorphic with U, and whose core is isomorphic with J, clearly the frame of S is isomorphic with S_0. This S_0 is a special case of the following type.

Associate with each element u of U a set $J(u)$ in any way subject to the following conditions:

(1) the sets $J(u)$ are mutually disjoint;

(2) $J(u)$ contains the element u of U;

(3) $J(z)$ consists of the given semigroup J, with the zero element of J identified with the identity element z of U.

For each $u \varepsilon U$ determine a right and a left representation of J by single-valued mappings of the set $J(u)$ into itself, subject to the following conditions:

(a) every left mapping commutes with every right mapping;

(b) the element z of J maps every element of $J(u)$ into u;

(c) all the mappings leave fixed the element u of $J(u)$.

We now proceed to define multiplication in the class sum S of all the sets $J(u)$. Within $J(z)$ it is defined as originally given in J. For an element k of J and an element a of $J(u)$, we define $ak(ka)$ to be the image of a under the mapping corresponding to k in the right (left) representation already determined. For $a_1 \varepsilon J(u_1)$ and $a_2 \varepsilon J(u_2)$, where $u_1 \neq z \neq u_2$, we define $a_1 a_2 = u_1 u_2$, where $u_1 u_2$ is the product of u_1 and u_2 as originally given in U.

The associativity conditions

$$(ak_1)k_2 = a(k_1 k_2) \text{ and } (k_1 k_2)a = k_1(k_2 a)$$

are the defining conditions for a right and a left representation, respectively, while

$$(k_1 a)k_2 = k_1(ak_2)$$

is just the commutation condition (a). Hence associativity holds in all cases

where two of the three factors are in J. From condition (c) and the above definition for a product a_1a_2 of elements not in J we have

$$k(a_1a_2) = k(u_1u_2) = u_1u_2,$$
$$(ka_1)a_2 = u_1u_2 \text{ (since } ka_1 \varepsilon J(u_1)).$$

Hence associativity holds if one of the three factors is in J. Finally, if none of the factors is in J, it follows from associativity in U.

Since U is clearly both a two-sided ideal and a group, it must be the group of zeroids of S. Since $J(z)$ consists of all elements of S for which z is a zero, it must be the core of S.

5. Subgroups and subsemigroups of a semigroup having zeroid elements. Throughout this section, S will be a semigroup having a non-vacuous group U of zeroid elements. As usual, we shall denote by z the identity element of U, and by $J(u)$ the congruence class of all elements x of S mapped into $u \varepsilon U$ by the homomorphism $\zeta : x \rightarrow zx = xz = u$. We shall denote by J alone the core $J(z)$ of S.

Let S' be a subsemigroup of S. Under ζ, S' maps into the subsemigroup zS' $(= S'z)$ of U. zS' may be characterized as the set of all $u' \varepsilon U$ such that $S' \cap J(u')$ is non-vacuous. Since zS' is part of a group, the cancellation law holds on both sides in zS', i. e. either $ab = ac$ or $ba = ca$ implies $b = c$. Thus S' is homomorphic with a semigroup in which cancellation holds, and consequently Dubreil's Theorems 29a and 29b (*l. c.*, p. 27) are applicable. (The reader should be warned that Dubreil uses the term "demi-group" for what we call "semigroup," and uses "semi-group" for a demi-group in which "simplification" (cancellation) holds on both sides. In the hypothesis of Theorem 29b, the author undoubtedly intended to assume that F is a semigroup, not just a demi-group as stated.)

LEMMA 5. *If G is a subgroup of S, then zG is a subgroup of U. $G \cap J$ is an invariant subgroup of G, and the factor group $G/G \cap J$ is isomorphic with zG. The cosets of $G \bmod G \cap J$ are precisely the intersections $G \cap J(u)$ of G with the congruence classes $J(u)$.*

Proof. This is an immediate consequence of the fundamental theorem of group homomorphisms applied to ζ.

LEMMA 6. *Every subgroup G of S lies either in U or in the complement of U.*

Proof. If $G \cap U$ is not vacuous, let v be an element thereof, and let g be any element of G. Since $v \varepsilon U$ there exists a $u \varepsilon U$ such that $gu = v$. Since $v \varepsilon G$ and G is a group, there exists a $g' \varepsilon G$ such that $vg' = g$. Hence $g = vg' = gug' \varepsilon U$, whence $G \subseteq U$.

THEOREM 4. *Let S' be a subsemigroup of S, and assume that S' itself has a non-vacuous group U' of zeroid elements. Then*

(1) *$zS' = zU'$, and consequently ζ maps S' onto the subgroup zU' of U;*

(2) *$U' \cap J$ is an invariant subgroup of U', and $U'/U' \cap J$ is isomorphic with zU';*

(3) *if a congruence class $J(u)$ contains an element of S' then it contains at least one zeroid element of S', and in fact $U' \cap J(u)$ is the coset of U'* mod *$U \cap J$ mapped into u by ζ;*

(4) *either $S' \cap U$ is vacuous, or else $U' \subseteq U$; in the latter event, if a congruence class $J(u)$ contains an element of S', then it contains exactly one zeroid element of S', namely u itself.*

Proof. (1) Let z' be the identity element of U'. Since z' is idempotent, it lies in $J = J(z)$. Hence $zz' = z$ and $zS' = zz'S' = zU'$.

(2) This follows from Lemma 5 since U' is a subgroup of S.

(3) The first part is a restatement of $zS' = zU'$, and the second follows from Lemma 5.

(4) If $S' \cap U$ is not vacuous, let $u' \varepsilon S' \cap U$. Then $z'u' \varepsilon U' \cap U$, so that $U' \cap U$ is not vacuous. By Lemma 6, $U' \subseteq U$. The second part is clear from (3) and the fact that $U \cap J(u)$ consists of u alone.

Since S' in Theorem 4 is homomorphic with the group $zS' = zU'$, Dubreil's Theorem 29c is applicable (*l. c.*, p. 28).

6. An embedding theorem. The following simple principle is sometimes useful in constructing semigroups.

LEMMA 7. *If S_1 and S_2 are disjoint semigroups, and if $a \rightarrow \bar{a}$ is a given homomorphism of S_1 into S_2, then the class sum S of S_1 and S_2 can be made into a semigroup containing S_1 and S_2 as subsemigroups by defining products in S as follows:*

(a) *within S_1 and S_2, products are defined as originally given;*

(b) *the product* a_1a_2 *of an element* a_1 *of* S_1 *with an element* a_2 *of* S_2 *is defined to be the product* $\bar{a}_1a_2$ *as given in* S_2, *and similarly* a_2a_1 *is defined to be* $a_2\bar{a}_1$.

A proof is unnecessary since it is a special case of a construction used in a previous paper by one of us.[4] It is, of course, merely a matter of verifying associativity. For example, if $a_1 \varepsilon S_1$, $b_1 \varepsilon S_1$, and $c_2 \varepsilon S_2$, then

$$(a_1b_1)c_2 = (\overline{a_1b_1})c_2 = (\bar{a}_1\bar{b}_1)c_2 = \bar{a}_1(\bar{b}_1c_2) = a_1(b_1c_2).$$

Use is made here of associativity within S_2, and of the homomorphism property $\overline{a_1b_1} = \bar{a}_1\bar{b}_1$.

THEOREM 5. *Let* U *be a group and* S' *a semigroup disjoint from* U. *Let* $a \rightarrow \bar{a}$ *be a homomorphism of* S' *into* U. *Then* S' *can be embedded in a semigroup* S *having* U *for its group of zeroid elements, in such a way that the given homomorphism coincides within* S' *with the homomorphism* ζ *of* S *onto* U.

Proof. Using Lemma 7, we merely take S to be the class sum of S' and U, and define

$$au \equiv \bar{a}u; \quad ua \equiv u\bar{a} \qquad (a \varepsilon S'; \quad u \varepsilon U).$$

Since U is a two-sided ideal in S, and at the same time a subgroup, it must be the group of zeroids of S. If $a \varepsilon S'$ and z is the identity element of U, then $za = z\bar{a} = \bar{a}$ and $az = \bar{a}z = \bar{a}$. Hence the given homomorphism $a \rightarrow \bar{a}$ coincides with $\zeta : a \rightarrow za = az$ within S'.

THE JOHNS HOPKINS UNIVERSITY.
UNIVERSITY OF TENNESSEE.

[4] A. H. Clifford, "Semigroups admitting relative inverses," *Annals of Mathematics*, vol. 42 (1941), pp. 1037-1049, § 3. In the proof of associativity (top of p. 1044) there is no need to assume that the S_α are groups rather than merely semigroups.

A PROBLEM IN DIMENSION THEORY.*

By J. H. Roberts.

1. Introduction. Suppose X is a separable metric space of dimension n, and Y is a $(2n+1)$-dimensional euclidean cube. Let Y^X denote the space of all single-valued continuous transformations (= *mappings*) of X into subsets of Y, with metric given by the formula

$$\rho(f, g) = \operatorname*{Sup}_{x \varepsilon X} \delta[f(x), g(x)],$$

where δ is the metric in Y. It is known[1] that if X is *compact* then the homeomorphisms of X into Y constitute a dense G_δ set in Y^X. Even if X is not compact, Y^X *contains* a dense G_δ set of homeomorphisms. In their book "Dimension Theory,"[1] Hurewicz and Wallman raise the question as to whether the homeomorphisms constitute a G_δ set, for non-compact X. The present paper answers this question in the negative. In fact it is shown that the set of all homeomorphisms may be a highly arbitrary set.

2. Notation. Let C be the Cantor ternary set of real numbers on $[0,1]$; i.e., $x \varepsilon C$ if and only if $0 \leqq x \leqq 1$ and x can be expressed in the ternary system without the use of the digit 1. Let D denote the set of two numbers 0 and 1. Let $H = C \times D$ and $K = C \times C$ (Cartesian product spaces). Then H and K are subsets of the square with vertices $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$. Also $\dim K = 0$, and H is a compact subset of K.

3. Lemma. *Let A be a subset of C such that $C - A$ is dense in C. Then there exists a subset X_0 of H, and a Cantor set of mappings F in K^{X_0}, homeomorphic to C under a homeomorphism g, such that if $f \varepsilon F$ and $g(f) = x$, then f is a homeomorphism of X_0 into K if and only if $x \varepsilon A$.*

Proof. Let X_0 be the subset of H consisting of all (x, y) with $x \varepsilon C - A$. For $\bar{x} \varepsilon C$ we define a mapping $f_{\bar{x}}$ *of* H into K. Write $C - \bar{x} = C_{1\bar{x}} + C_{2\bar{x}} + \cdots$ where $C_{1\bar{x}}$ consists of all $x \varepsilon C$ on that one of the two intervals $0 \leqq x \leqq 1/3$, $2/3 \leqq x \leqq 1$ which does not contain $\bar{x}$; $C_{2\bar{x}}$ consists of all $x \varepsilon C$

* Received March 29, 1947; Presented to the American Mathematical Society, February 22, 1947.

[1] See Hurewicz and Wallman, *Dimension Theory*, Princeton University Press, Theorem V2, p. 56. See also Theorem V3 and the last footnote on p. 60.

on that one of the four intervals bounded by $[0, 1/9]$, $[2/9, 1/3]$, $[2/3, 7/9]$, $[8/9, 1]$ which does not contain $\bar{x}$ and is not a subset of $C_{1\bar{x}}$; etc. For $x \varepsilon C_{n\bar{x}}$ set $f_{\bar{x}}(x, y) = (x, y/3^n)$, for $(x, y) \varepsilon H$. For $x = \bar{x}$ set $f_{\bar{x}}(x, y) = (x, 0)$. Then since y is 0 or 1 it follows that $y/3^n \varepsilon C$ so that $f_{\bar{x}}$ is a mapping of H into K. Now consider the mapping $f_{\bar{x}}$ as applying only to X_0, and let F be the collection of all such $f_{\bar{x}}$ for $\bar{x} \varepsilon C$.

3.1. *The correspondence $g(f_{\bar{x}}) = \bar{x}$ is a homeomorphism.*

(a) g is $(1-1)$; i.e., if $x_1 \neq x_2$ then f_{x_1} is not identical with f_{x_2} over X_0. This follows from the definition of $f_{\bar{x}}$ and the facts that for some n, $C_{nx_1} \cdot C_{nx_2} = 0$, and $C - A$ is dense in C.

(b) g is uniformly continuous. For suppose $| x_1 - x_2 | < 1/3^k$, k a positive integer, and $(x, y) \varepsilon H$. Let y_i be the y-coordinate of $f_{x_i}(x, y)$, $(i = 1, 2)$. Then $\rho(f_{x_1}(x, y), f_{x_2}(x, y)) = | y_1 - y_2 |$. Now $C_{mx_1} = C_{mx_2}$ for $m = 1, 2, \cdots, k$. Hence $y_1 = y_2$, or else $y_i < 1/3^k$ $(i = 1, 3)$. Thus in any case $\rho(f_{x_1}, f_{x_2}) < 1/3^k$.

(c) g^{-1} is continuous. This follows from (a) and (b), since F is compact. This completes the proof of 3.1.

3.2. If $g(f) = \bar{x}$, $f \varepsilon F$, *then f is a homeomorphism of X_0 into K if and only if $\bar{x} \varepsilon A$.*

Now $f = f_{\bar{x}}$. If $\bar{x} \not\varepsilon A$ then $\bar{x} \varepsilon C - A$ so that $(\bar{x}, 0) \varepsilon X_0$ and $(\bar{x}, 1) \varepsilon X_0$. But $f_{\bar{x}}(\bar{x}, 0) = f_{\bar{x}}(\bar{x}, 1) = (\bar{x}, 0)$, so that $f_{\bar{x}}$ is not $1-1$; hence it is not a homeomorphism.

If $\bar{x} \varepsilon A$, then $(\bar{x}, y) \not\varepsilon X_0$ for any y. Then X_0 is the sum of the disjoint sets $S_1, S_2, \cdots$, where S_n is the set of all $(x, y) \varepsilon X_0$ with $x \varepsilon C_{n\bar{x}}$. It is obvious from the definition that $f_{\bar{x}}$ is a homeomorphism over each S_n, and each S_n is both open and closed in X_0. Hence $f_{\bar{x}}$ is a homeomorphism over X_0. This completes the proof of the lemma.

4. THEOREM. *Suppose $n \geqq 0$, and A is a subset of a Cantor set C such that $C - A$ is dense in C. Let Y denote an $(n+1)$-dimensional euclidean cube. Then there exists an n-dimensional set X with the following property: In the space Y^X there is a set T, corresponding to C under a homeomorphism g, such that if $t \varepsilon T$ and $g(t) = x$, $(x \varepsilon C)$, then t is a homeomorphism of x into Y if and only if $x \varepsilon A$.*

Proof. For $n = 0$, we use the lemma. Let h be any homeomorphism of the 0-dimensional set K into an interval Y. Let T be the set of all

mappings hf, $f \varepsilon F$. Then hf is a homeomorphism of X into Y if and only if f is a homeomorphism of X_0 into K.

For $n > 0$ we define X to be the Cartesian product of X_0 by n intervals, $(0 \leqq z \leqq 1)$. Then if $p \varepsilon X$, $p = (x, y, z_1, z_2, \cdots, z_n)$, where $(x, y) \varepsilon X_0$ and $0 \leqq z_i \leqq 1$ for $i = 1, 2, \cdots, n$. Consider Y as the set of all points $(z_0, z_1, \cdots, z_n)$, $0 \leqq z_i \leqq 1$, $i = 0, 1, \cdots, n$. Then corresponding to an $f \varepsilon F$ define $t(X) \subset Y$ as follows:

$$t(x, y, z_1, z_2, \cdots, z_n) = (hf(x, y), z_1, z_2, \cdots, z_n).$$

The collection T of all such t meets the requirements of the theorem.

Conclusion. To solve the originally posed problem it is sufficient to take A as a non-G_δ subset of C; e. g., any dense countable subset of C. Then the set of all homeomorphisms of X into Y cannot be a G_δ set in Y^X, since the intersection of a G_δ set and the compact set T would be a G_δ in T. The set Y is an $(n + 1)$-cube, hence is a subset of a $(2n + 1)$-cube. We can state the following final result:

If $n \geqq 0$, $k \geqq n + 1$, and Y is a k-dimensional cube, then there exists an n-dimensional subset X of an $(n + 1)$-cube such that in the space Y^X the homeomorphisms do not constitute a G_δ set.

DUKE UNIVERSITY.

A KINEMATIC CHARACTERIZATION OF SERIES OF LINEAL ELEMENTS IN THE PLANE AND OF THEIR DIFFERENTIAL INVARIANTS UNDER THE GROUP OF WHIRL-SIMILITUDES AND SOME OF ITS SUBGROUPS.*

By J. M. Feld.

1. Introduction. Let T_α represent the transformation that rotates every oriented lineal element in the plane through the same angle α about its point. Let S_k represent the transformation that slides every oriented lineal element a distance k along its line. The product $T_\alpha S_k T_\beta$ is called a *whirl*—a term introduced by Kasner[1] [9]. The whirls constitute a three-parameter group of lineal element transformations G_3. The product of a whirl and a euclidean displacement, in Kasner's terminology, is a *whirl-motion*. The whirl-motions constitute a six-parameter group G_6. The geometry of whirl-motions in the plane was first studied by Kasner [9], and has since been developed in a series of papers by Kasner, DeCicco, and the author.[2] The product of a whirl by direct similtude will be called a *whirl-similitude*. The whirl-similitudes[3] form a seven-parameter group G_7 [7].

DeCicco [4] has found a set of fundamental differential invariants of series of lineal elements under G_6. In this paper we determine differential invariants of series under G_3, G_6, and G_7. DeCicco, in his investigation, regarded a series in the plane as an analogue of a curve in space, and was guided by consideration of the properties of *euclidean* curvature and torsion of curves in his definitions of the "curvature" and "torsion" of series. However, it was shown by the author[4] that the whirl-motion group in the

* Received May 18, 1947; Presented to the American Mathematical Society, September 2, 1947.

[1] The numbers in brackets refer to the bibliography at the end of the paper.

[2] A geometry of whirl-motions on the sphere was investigated by Strubecker [13]. Generalizations of whirl-motions to higher spaces were made by Narasinga Rao [12] and by Feld [7].

[3] G_7 is a continuous subgroup of a mixed group Γ_7 composed of eight continuous families of lineal element transformations [6].

[4] In [6, 8] we showed how lineal elements in the plane can be mapped (1, 1) and continuously upon the points of *quasi-elliptic* three-space Q_3 in such a way that the group of whirl-motions G_6 is simply isomorphic to the group of quasi-elliptic motions $\mathfrak{G}_6$ in Q_3. The geometry of Q_3 under $\mathfrak{G}_6$ has been investigated by Blaschke [1, 2]; many

plane is isomorphic to the group of motions in *quasi-elliptic three-space* Q_3, and not to the group of motions in euclidean three-space. Consequently, DeCicco's procedure yields invariants having rather complicated geometric interpretations. A different point of view is adopted in this paper: we regard a series of lineal elements as a description of the ∞^1 positions that a plane takes when it is subjected to a continuous displacement over another plane. This point of view leads to three sets of fundamental differential invariants, one for each of the groups G_3, G_6, and G_7, which can be given simple kinematic characterizations.

2. Differential invariants of series under the group of whirls G_3. Let the point of the oriented lineal element e have the Cartesian coordinates x, y; let the angle of inclination of e to an arbitrary given direction be θ. A series S is then given by the equations

$$x = f_1(t), \qquad y = f_2(t), \qquad \theta = f_3(t)$$

where the functions $f_i(t)$ are assumed to have continuous derivatives of the orders required by the conditions that arise. If $f'_3(t) \equiv 0$, S is a series of parallel lineal elements (*parallel series*). Unless otherwise stated, the term *series* shall mean only non-parallel series; parallel series will be treated separately. If $f'_3(t) \neq 0$ for $t_1 \leqq t \leqq t_2$, then, in this interval S can be represented by the equations $x = g_1(\theta)$, $y = g_2(\theta)$. Letting $z = x + iy$, $\bar{z} = x - iy$, a series S can be represented by

$$z = f(\theta). \tag{2.1}$$

We asume that $f'(\theta) \neq 0$ for $\theta_1 \leqq \theta \leqq \theta_2$. We shall call z, θ the Gaussian coordinates [5] of e.

The locus of the points of the elements of S will be called the point locus of S and the envelope of the lines of the elements of S its line locus. Parallel series have no line locus. The series defined by

$$z = l + re^{i\theta} \tag{2.2}$$

has a circle of radius $|r|$ and center l for its point locus and a concentric

of Blaschke's results are therefore in the abstract equivalent to corresponding discoveries made by Kasner and DeCicco. In particular, Blaschke has determined a set of fundamental differential invariants for curves under $\mathfrak{G}_6$.

[5] We introduced these coordinates in [6]. Their use materially simplifies the computation. Kasner and DeCicco use Hessian and Cartesian coordinates.

oriented circle for its line locus. Such a series has been named a *turbine* by Kasner. Let l, r be called the coordinates of the turbine.

As was shown in [6], the group of whirls in Gaussian coordinates is given by the transformations

$$(2.3) \qquad z^* = z + be^{i\theta}, \qquad \theta^* = \theta + \alpha, \qquad (\alpha \text{ real}).$$

Turbines l, r and l^*, r^* such that $l = l^*$ can be transformed into each other by suitable whirls. A series S given by (2.1), when subjected to a whirl, is transformed into a series S^* having for its equation

$$z^* = f(\theta^* - \alpha) + be^{i(\theta^* - \alpha)}.$$

Two such series, S and S^*, will be called equivalent under G_3 (within a whirl). We shall find necessary and sufficient conditions for the equivalence of two series under G_3.

Equations (2.3) yield $d\theta^* = d\theta$ and $z^* + i(dz^*/d\theta^*) = z + i(dz/d\theta) \equiv I(\theta)$. In accordance with what is now well established usage, $d\theta$ will be called the differential of G_3-*arc length*, and the real and imaginary parts of the differential invariant $I(\theta)$, namely $x(\theta) - y'(\theta)$ and $x'(\theta) + y(\theta)$, will be called the *first* and *second curvatures* respectively of S. The equation

$$(2.4) \qquad I = F(\theta)$$

determines, as will presently be proved, a series within a whirl. Consequently, (2.4) may be regarded as the intrinsic equation of a series relative to G_3.

Let $z = f(\theta)$ be a particular solution of

$$(2.4^*) \qquad I \equiv z + iz' = F(\theta).$$

Then the general solution of (2.4*) is $z^* = f(\theta) + be^{i\theta}$. Since the G_3-arc length θ may be measured from any initial direction, we can set $\theta^* = \theta + \alpha$, and write the general solution in the form $z^* = f(\theta^* - \alpha) + be^{i(\theta^* - \alpha)}$, which is the transform of $z = f(\theta)$ by a whirl.

The series for which $I = l$ (constant) are the turbines $z = l + re^{i\theta}$ having a common center at the point $z = l$.

Series of lineal elements and their differential invariants can be interpreted kinematically as follows:

Let a plane π_1 be displaced continuously over a fixed plane π. Let $\mathfrak{e}_1$ be a fixed (primitive) oriented lineal element on π_1. As π_1 is displaced, $\mathfrak{e}_1$ traces out on π a series of lineal elements. This series completely defines

the motion of π_1 on π. If the motion of π_1 is a translation, the corresponding series is a parallel series, the point locus of which is traced out by the point of $\mathfrak{e}_1$. If the motion is not a translation, the corresponding series can be represented by an equation: $z = f(\theta)$, $(z' \neq 0)$, where z, θ are the coordinates of the primitive element $\mathfrak{e}_1$ on π_1 referred to rectangular axes on π; or, what is the same thing, $\mathfrak{e}_1$ may be referred to a primitive element $\mathfrak{e}$ on π, where the point of $\mathfrak{e}$ is at the origin and the direction of $\mathfrak{e}$ is the positive direction of the x-axis. A whirl can now be interpreted as the replacement of $\mathfrak{e}_1$ on π_1 by another primitive element $\mathfrak{e}^*_1$. While $\mathfrak{e}_1$ describes the series S, $\mathfrak{e}^*_1$ describes the series S^*. Two series S and S^* determine the same continuous motion of π_1 on π if and only if they are equivalent under the group of whirls.

Let w be the Gaussian coordinate of a point P on π_1 referred to the primitive element $\mathfrak{e}_1$; then the Gaussian coordinate of P referred to the primitive element $\mathfrak{e}$ on π is $z_1 = z + we^{i\theta}$ and the point locus of z_1 is the trajectory that P traces on π in the course of the motion defined by (2.1). The points $z_1(\theta)$, such that $z'_1(\theta) = 0$, are the poles (instantaneous centers of rotation) on π. The locus of these poles on π is the fixed (space) centrode of the motion, and the locus of the poles on π_1 is the mobile (body) centrode. The fixed centrode (referred to $\mathfrak{e}$) is given by the point locus of

$$Z = z + iz', \tag{2.5}$$

and the mobile centrode (referred to $\mathfrak{e}_1$) is given by the point locus of

$$W = iz'e^{-i\theta}. \tag{2.6}$$

It is well known that a continuous non-translatory motion (2.1) is equivalent to the rolling without slipping of a curve W in plane π_1 upon a curve Z in plane π. The point locus of the complex invariant $I(\theta)$ is the fixed centrode of the motion determined by the series (2.1). If $I(\theta) = \text{constant}$, the fixed centrode, and therefore also the mobile centrode, degenerates to a point; the corresponding motion is the rotation of the mobile plane around a fixed point in the fixed plane.

A turbine is determined by two non-parallel lineal elements. A turbine determined by two adjacent lineal elements of the series (2.1) is said to be tangent to the series. The coordinates of the tangent turbine at z, θ are given by $l = z + iz'$, $r = -iz'e^{-i\theta}$. Hence, we find that the fixed and mobile centrodes of a motion defined by a series are given by the point loci of l and $-r$, respectively, where l, r are the coordinates of the ∞^1 turbines tangent to the series.

If the mobile primitive element $\mathfrak{e}_1$ has the coordinates z, θ when referred to the fixed primitive element $\mathfrak{e}$, then $\mathfrak{e}$ has, when referred to $\mathfrak{e}_1$, the coordinates z^*, θ^*, where

$$z^* = -ze^{-i\theta}, \qquad \theta^* = -\theta. \tag{2.7}$$

Let $\mathfrak{M}$ denote the notion of π_1 on π defined by the series (2.1); then the reverse motion $\mathfrak{M}^{-1}$ of π on π_1 is given by the transform of (2.1) by the involutory transformation (2.7). Thus $\mathfrak{M}^{-1}$ denotes the motion given by the series

$$z^* = -f(-\theta^*)e^{i\theta^*}. \tag{2.1*}$$

The fixed and mobile centrodes of $\mathfrak{M}^{-1}$ are respectively the mobile and fixed centrodes of $\mathfrak{M}$.

If π_1 is subjected to a translatory motion over π, two lineal elements $\mathfrak{e}_1$ and $\mathfrak{e}'_1$ in π_1 trace out two parallel series S and S' respectively. Let $\mathfrak{T}$ be the translation that carries the point of $\mathfrak{e}_1$ into the point of $\mathfrak{e}'_1$. Then $\mathfrak{T}$ will transform the point locus of S into the point locus of S'. Two parallel series are equivalent under G_3 if and only if the point locus of one is the transform by a translation of the other.

3. Differential invariants under the group of whirl-similitudes. The seven-parameter group of whirl-similitudes G_7 is given [6] by the equations

$$z^* = az + be^{i\theta} + c, \qquad \theta^* = \theta + \alpha, \qquad (\alpha \text{ real}). \tag{3.1}$$

Under G_7, as under G_3, $d\theta$ is invariant, and will be designated as the differential of G_7-arc length of the series $S : z = f(\theta)$. Evidently

$$\frac{d^n z^*}{d\theta^{*n}} + i\frac{d^{n+1}z^*}{d\theta^{*n+1}} = a\left(\frac{d^n z}{d\theta^n} + i\frac{d^{n+1}z}{d\theta^{n+1}}\right), \qquad n \geqq 1, \tag{3.2}$$

yielding, consequently, as complex differential invariant of lowest order for S

$$J(\theta) \equiv (z'' + iz''')/(z' + iz''). \tag{3.3}$$

We assume that S has a continuous third order derivative, and since turbines have no invariants under G_7, we assume furthermore that S is not a turbine, that is, that $z' + iz'' \not\equiv 0$.

Let the real and imaginary parts of $Z(\theta) = z + iz'$ be $u(\theta)$ and $v(\theta)$ respectively. Hence

$$J(\theta) = Z''/Z' = \frac{(u'u'' + v'v'') + i(u'v'' - v'u'')}{[(u')^2 + (v')^2]^2}.$$

Let s be the euclidean arc length of the fixed centrode $Z(\theta)$ measured from an arbitrary point θ_0 on the centrode to the point θ, and let ω be the angle of inclination of the tangent to the centrode at the point θ; then the real and imaginary parts of $J(\theta)$ are respectively equal to

$$(d^2s/d\theta^2)/(ds/d\theta) \quad \text{and} \quad (d\omega/d\theta).$$

These we shall call the first and second curvatures respectively of S under G_7. If we regard θ as a function of the time t, $(ds/d\theta) = (ds/dt)/(d\theta/dt)$ where ds/dt denotes the rate of displacement of the pole on the fixed centrode and $d\theta/dt$ denotes the angular rate of rotation of the mobile plane. Likewise, $d\omega/d\theta(= (d\omega/dt)/(d\theta/dt))$ denotes the ratio of the rate of turning of the tangent to the centrode to the rate of turning of the mobile plane.

We shall prove that $J(\theta)$, that is, the pair of G_7-curvatures, determines a series within a whirl-similitude, and that consequently an equation such as

$$J = F(\theta) \tag{3.4}$$

can be regarded as the intrinsic equation of a series under G_7.

Equation (3.4), by virtue of (3.3), yields

$$z''' - (F + i)z'' + iFz' = 0. \tag{3.5}$$

Evidently (3.5) is satisfied by $z = c$ (constant) and by $z = e^{i\theta}$. Let $z = f(\theta)$ be a particular solution of (3.5) linearly independent of c and $e^{i\theta}$; then the general solution of (3.5) is $z^* = af(\theta) + be^{i\theta} + c$. Since θ is measured from an arbitrary direction, we can set $\theta^* = \theta + \alpha$ and express the general solution in the form $z^* = af(\theta^* - \alpha) + be^{i(\theta^* - \alpha)} + c$, which is the transform of $z = f(\theta)$ by a whirl-similitude.

The series that have constant G_7-curvatures have for their intrinsic equation

$$J(\theta) = k \text{ (constant)}. \tag{3.6}$$

If $k \neq 0, i$, the series determined by (3.6) are of type

$$z = ae^{k\theta} + be^{i\theta} + c. \tag{A}$$

If $k = i$, the series are of type

$$z = (a + b\theta)e^{i\theta} + c. \tag{B}$$

If $k = 0$, the series are of type

$$z = a\theta + be^{i\theta} + c. \tag{C}$$

MOTIONS OF TYPE (A): Within a whirl-similitude the series of this type are equivalent to $z = e^{(k_1+ik_2)\theta}$ where $k = k_1 + ik_2$. The corresponding motion has for its fixed centrode

$$Z = (1 - k_2 + ik_1)\, e^{(k_1+ik_2)\theta}$$

and for its mobile centrode

$$W = (-k_2 + ik_1)\, e^{[k_1+i(k_2-1)]\theta}$$

(A_1): Let the first curvature $k_1 = 0$. Since the second curvature $k_2 \neq 0, 1$, the fixed centrode is a circle of radius $|1 - k_2|$ and the mobile centrode is a circle of radius $|k_2|$. The motion is therefore that of a circle rolling without slipping on another circle.

(A_2): Let $k_2 = 0$; then $k_1 \neq 0$. Now the fixed centrode is evidently the line $y = k_1 x$. Letting s and ρ denote respectively the euclidean arc length and radius of curvature of the mobile centrode, we find that its euclidean intrinsic equation is $\rho = k_1 s$, namely that of an equiangular spiral.

(A_3): Let $k_2 \neq 0, 1$, and let $k_1 \neq 0$. The fixed and mobile centrodes have as their euclidean intrinsic equations, respectively, $k_2\rho = k_1 s$ and $(k_2 - 1)\rho = k_1 s$. The motion is therefore that of an equiangular spiral rolling on another equiangular spiral.

(A_4): Let $k_2 = 1$ and let $k_1 \neq 0$. By virtue of equation (2.1*), the motion defined in this case is seen to be the inverse of that described in (A_2); consequently, it is that of a line rolling on an equiangular spiral.

MOTION OF TYPE (B). Here $k_1 = 0$ and $k_2 = 1$. Within a whirl-similitude the series of type (B) are equivalent to $z = \theta e^{i\theta}$. The corresponding motion has for its fixed centrode $Z = ie^{i\theta}$ and for its mobile centrode $W = i - \theta$. This motion is evidently that of a line rolling on a circle.

MOTION OF TYPE (C). $k_1 = k_2 = 0$. The motion defined by series of this type is inverse to that of type (B); consequently, it is that of a circle rolling on a line.

A parallel series S is transformed by a whirl-similitude into a parallel series S^*. Let S and S^* be two parallel series and let C and C^* be their respective point loci. Then, a necessary and sufficient condition that $S \rightarrow S^*$ by means of a whirl-similitude is that $C \rightarrow C^*$ by means of a direct similitude.

4. Whirl-motion invariants. The six-parameter group of whirl-motions G_6 is defined by equations (3. 1) when $|a|=1$. Since G_6 is transitive when applied to turbines, turbines have no invariants under G_6. We shall therefore consider only differential invariants of series $z(\theta)$ such that $z'+iz''\neq 0$.

Once again $d\theta$ is the differential of G_6-arc length. Since $|a|=1$, we obtain from (3. 2) the differential invariants

$$(4.1)\qquad (z'+iz'')(\bar{z}'-i\bar{z}'')$$

and

$$(4.2)\qquad (z''+iz''')/(z'+iz'').$$

With $Z=z+iz'=u(\theta)+iv(\theta)$, equation (4. 1) takes the form

$$(4.1^*)\qquad Z'\bar{Z}'=(u')^2+(v')^2=(ds/d\theta)^2$$

where s is the euclidean arc length of the fixed centrode $Z(\theta)$. We have seen that (4. 2) implies the two real invariants $(d^2s/d\theta^2)/(ds/d\theta)$ and $(d\omega/d\theta)$. Since $ds/d\theta$ is invariant, so is $d^2s/d\theta^2$. Therefore, (4. 2) yields only one other fundamental invariant, namely $d\omega/d\theta$. We shall call $\kappa_1=ds/d\theta$ and $\kappa_2=d\omega/d\theta$ the first and second curvatures respectively under G_6. Evidently κ_2/κ_1, which is the euclidean curvature of the fixed centrode, is also invariant. The two curvatures κ_1 and κ_2 determine, as will be proved presently, a series within a whirl-motion; we shall consequently regard

$$(4.3)\qquad \kappa_1=\beta(\theta)\neq 0,\qquad \kappa_2=\gamma(\theta)$$

as the intrinsic equations of a series (other than turbines) under G_6. We assume that $\beta(\theta)\neq 0$, $\theta_1\leqq\theta\leqq\theta_2$, and that it is a real function possessing a continuous first derivative; also that $\gamma(\theta)$ is a continuous real function. From (4. 3) we obtain

$$Z''/Z'=\beta'/\beta+i\gamma;$$

hence

$$(4.4)\qquad Z'=z'+iz''=c_1\beta e^{i\int\gamma d\theta}.$$

Furthermore, since

$$(4.5)\qquad Z'\bar{Z}'=(z'+iz'')(\bar{z}'-i\bar{z}'')=\beta^2,$$

$|c_1|=1$. Thus a series determined by the intrinsic equations (4. 3) must satisfy (4. 4) and (4. 5). Let $z=f(\theta)$ be a particular solution of these equations. Then we obtain as the general solution $z^*=af(\theta)+be^{i\theta}+c$, $|a|=1$, or if we let $\theta^*=\theta+\alpha$

$$z^* = af(\theta^* - \alpha) + be^{i(\theta^* - \alpha)} + c, \qquad |a| = 1, \tag{4.6}$$

which is the transform of $z = f(\theta)$ by a whirl-motion.

Let us determine the series for which the G_6-curvatures κ_1 and κ_2 are constant. If $\kappa_1 = 0$, $(z' + iz'' = 0)$, the series determined are the ∞^4 turbines $z = l + re^{i\theta}$. If $\kappa_1 \neq 0$, we get three families of series, namely:

A: $\kappa_1 = k_1 \neq 0, \qquad \kappa_2 = k_2 \neq 0, 1;$

$$z = \frac{c_1 k_1}{k_2(1 - k_2)} e^{ik_2\theta} + c_2 e^{i\theta} + c_3, \qquad |c_1| = 1.$$

B: $\kappa_1 = k_1 \neq 0, \qquad \kappa_2 = 0;$

$$z = c_1 k_1 \theta + c_2 e^{i\theta} + c_3, \qquad |c_1| = 1.$$

C: $\kappa_1 = k_1 \neq 0, \qquad \kappa_2 = 1;$

$$z = c_1 k_1 \theta e^{i\theta} + c_2 e^{i\theta} + c_3, \qquad |c_1| = 1.$$

The series in family A are equivalent within a whirl-motion to the series

$$z = (k_1/k_2(1 - k_2))e^{ik_2\theta}. \tag{4.7}$$

The motion defined by (4.7) has for its fixed centrode the circle $Z = (k_1/k_2)e^{ik_2\theta}$, and for its mobile centrode the circle

$$W = [k_2/(k_2 - 1)]e^{i(k_2-1)\theta}.$$

The series in family B are equivalent, within a whirl-motion, to the series

$$z = k_1\theta, \tag{4.8}$$

which has for its fixed centrode the line $Z = k_1\theta + ik_1$, and for its mobile centrode the circle $W = ik_1 e^{-i\theta}$.

The series in family C determine motions which are the inverses of those determined by the series in family B; consequently, their fixed centrodes are circles and their mobile centrodes lines.

Whirl-motions convert parallel series into parallel series. Let S and S^* be two parallel series and let C and C^* be their respective point loci; then a necessary and sufficient condition that $S \to S^*$ by means of a whirl-motion is that $C \to C^*$ by means of a euclidean displacement.

QUEENS COLLEGE,
FLUSHING, N. Y.

BIBLIOGRAPHY

1. W. Blaschke, "Euklidische Kinematik und nichteuklidische Geometrie," *Zeitschrift für Mathematik und Physik*, vol. 60 (1911), pp. 61-91 and 203-204.

2. ———, *Ebene Kinematik*, Leipzig and Berlin, 1938.

3. J. DeCicco, "The geometry of whirl series," *Transactions of the American Mathematical Society*, vol. 43 (1938), pp. 344-358.

4. ———, "The differential geometry of series of lineal elements," *Transactions of the American Mathematical Society*, vol. 46 (1939), pp. 348-361.

5. ———, "The geometry of fields of lineal elements," *Transactions of the American Mathematical Society*, vol. 47 (1940), pp. 207-229.

6. J. M. Feld, "Whirl-similitudes, euclidean kinematics, and non-euclidean geometry," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 783-790.

7. ———, "The geometry of whirls and whirl-motions in space," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 927-933.

8. ———, "On a representation in space of groups of circle and turbine transformations in the plane," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 930-934.

9. E. Kasner, "Turns, slides and the geometry of turbines," *American Journal of Mathematics*, vol. 33 (1911), pp. 193-202.

10. Kasner and DeCicco, "Geometry of turbines, flat fields, and differential equations," *American Journal of Mathematics*, vol. 59 (1937), pp. 545-563.

11. ———, "The geometry of the whirl-motion group G_6: elementary invariants," *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 399-403.

12. A. Narasinga Rao, "Studies in turbine geometry I," *Journal of the Indian Mathematical Society*, vol. 3 (1938), pp. 96-108; II, *Proceedings of the Indian Academy of Sciences*, vol. 8A (1938), pp. 179-186.

13. K. Strubecker, "Zur Geometrie sphärischer Kurvenscharen," *Jahresbericht der deutschen Mathematiker-Vereinigung*, vol. 44 (1934), pp. 184-198.

BOOLEAN ALGEBRA IN TERMS OF INCLUSION.*

By LEE BYRNE.

Ordinary axiomatizations of Boolean algebra may be separated into two classes, according as their original, indefined concepts are (1) *operations* (functions) such as product, sum, complement (employed along with a relation of equality or identity), or (2) *a relation* of inclusion; while, of course, some combination of the two is also possible. Examples of the operational type have been much the more numerous. Moreover, all axiomatizations in a small number of postulates have been of the operational type, while for the relational type the tendency has been for the number of postulates to be much larger.

From the present paper it will become evident that a relational formulation can in fact show a comparable economy as regards number of postulates. There may be a further interest in seeing how the development of the system might proceed in purely relational terms.[1]

Four axioms will be used, which cover the ground commonly assigned to "transformation" postulates, and in addition those assumptions as to closure, number of elements, and properties of an equality relation, which are often made only tacitly or informally.[2] Enough theorems will be proved to show that the system is a Boolean algebra. This result will be obtained through comparison with the postulates known as Huntington's Second Set.

Aside from the class of elements in the system (which is not explicitly symbolized), the present formulation uses no undefined concept except a dyadic relation of inclusion, symbolized by $\subset$, the character of which is determined only by the axioms.

To prevent theorems and even axioms from reaching inconvenient length, six concepts are introduced by initial definition, two being for properties, two for dyadic relations, and two for triadic relations. No concepts of operation (function) are introduced.

* Received May 8, 1947.

[1] In previous relational formulations it has been customary to make early definitional introduction of the usual operations, and then to proceed with the development very much as would be done in a version that was operational from the outset.

[2] Thanks are due to referee for suggestions leading to shorter and simpler proofs and influencing the final choice of axioms.

Lower case italic letters, with or without accents, are used as variables for elements in the system.

In order to form our statements about the elements and relations of the system, we shall also employ certain symbols which may be regarded as taken over from the language of mathematical logic, viz.: $\sim$ 'not' as negative of a sentence, & 'and' as a conjunction between sentences, $\rightarrow$ 'if . . . then' or 'only if' for (material) implication between sentences, $\leftrightarrows$ 'if and only if' for (material) equivalence between sentences, (x) 'for every x' as universal operator (quantifier) on sentence forms, and $(\exists x)$ 'there exists an x such that' as existential operator (quantifier) on sentence forms. No parentheses are required around arguments of relations or properties, but parentheses are used in customary ways around sentence parts, whenever necessary.

Definitions

AN. $Nx \leftrightarrows (y)(x \subset y)$.

AU. $Uy \leftrightarrows (x)(x \subset y)$.

AC. $Cyz \leftrightarrows (x)(x \subset y \& x \subset z \rightarrow Nx) \& (w)(y \subset w \& z \subset w \rightarrow Uw)$.

AP. $Pyzw \leftrightarrows y \subset z \& y \subset w \& (x)(x \subset z \& x \subset w \rightarrow x \subset y)$.

AS. $Sxzw \leftrightarrows z \subset x \& w \subset x \& (y)(z \subset y \& w \subset y \rightarrow x \subset y)$. (Not used in paper).

AE. $x \supset \subset y \leftrightarrows x \subset y \& y \subset x$.

After Theorem 2.9:

DE. $x = y \leftrightarrows x \supset \subset y$.

The definitions connect the definiendum and definiens by means of the ordinary sign of equivalence between sentences, and are all "definitions in use." A general label 'A' is used to suggest their character as abbreviations. With the aid of these definitions the following axioms can be expressed with relative brevity.

Axioms

I(a). $(\exists x)(\exists y)(\sim(x \subset y))$.

I(b). $(x)(y)(\exists p)(\exists p')(Ppxy \& Cpp')$.

II(a) $Cyy' \& Ppxy' \& Np \rightarrow x \subset y$.

II(b) $x \subset y \& y \subset z \rightarrow x \subset z$.

The theorems which follow are numbered with Arabic numerals. Steps in proof are labelled with a, b, c etc. After a step an indication is commonly given of the source from which it is derived, by a small letter if another step, AN etc. if a definition, a Roman numeral if an axiom, an Arabic numeral if another theorem.

Theorems

1. $x \subset x$.

Proof. a. $(x)(\exists p)(\exists p')(\exists q)(Ppxx \& Cpp' \& Pqxp')$. I(b).

b. $Ppxx \& Cpp' \& Pqxp' \rightarrow Nq$. AP, II(b), AC.

c. $(x)(\exists p)(x \subset p \& p \subset x)$. a, b, AP, II(a).

Hence theorem by II(b).

2.1. $x \supset\subset x$. AE, 1.

2.2. $x \supset\subset y \rightarrow y \supset\subset x$. AE.

2.3. $x \supset\subset y \& y \supset\subset z \rightarrow x \supset\subset z$. AE, II(b).

2.4. $x \supset\subset y \rightarrow (x \supset\subset z \leftrightarrows y \supset\subset z)$. AE, II(b).

2.5. $y \supset\subset z \rightarrow (x \supset\subset y \leftrightarrows x \supset\subset z)$. AE, II(b).

2.9. (metamathematical) The relation expressed through sentences of the form '$x \supset\subset y$' has all the properties customarily attributed to an equality relation '$x = y$.'[3]

3. $(x)(y)(\exists p)(Ppxy)$. I(b).

4. $(x)(\exists c)(Cxc)$.

Proof. a. $(x)(\exists p)(\exists c)(Ppxx \& Cpc)$. I(b).

b. $Ppxx \rightarrow (p \subset x \& x \subset p)$. AP, 1.

c. $p \subset x \& x \subset p \& Cpc \rightarrow Cxc$. AE, AC, 2.4, 2.5.

Hence theorem from b, c.

5. $(\exists n)(Nn)$.

Proof. a. $(x)(\exists x')(\exists n)(Cxx' \& Pnxx')$. 4, 3.

b. $Cxx' \& Pnxx' \rightarrow Nn$. AC, AP.

[3] In a fuller presentation we could successively show that the assertion holds for atomic, molecular, and general sentences, and hence for all sentence forms used in the system. The atomic sentences of the system are exclusively of the form '$x \subset y$'.

Hence theorem from a, b.

6. $Cxx'\&Cyy'\&x \subset y' \rightarrow y \subset x'$.

Proof. a. $(x)(y)(\exists p)(Ppxy)$. 3.

b. $Cyy'\&Ppxy \rightarrow (x \subset y' \rightarrow Np)$. AP, II(b), AC.

c. $Cxx'\&Ppyx \rightarrow (Np \rightarrow y \subset x')$. II(a), AC.

Hence theorem from a, b, c, AP.

7. $(\exists u)(Uu)$.

Proof. a. $(\exists n)(x)(\exists x')(Cxx'\&n \subset x')$. 5, 4, AN.

b. $Cxx'\&Cnn' \rightarrow (n \subset x' \rightarrow x \subset n')$. 6.

c. $(\exists n)(\exists n')(x)(\exists x')(Cxx'\&x \subset n')$. a, b, 4.

Hence theorem from c, AU.

8. $Cyy'\& \sim (x \subset y') \rightarrow (\exists a)(\sim Na\&a \subset x\&a \subset y)$.

Proof. a. $Cy'y \rightarrow (Ppxy\&Np \rightarrow x \subset y')$. II(a).

b. $Cyy' \rightarrow (\sim (x \subset y') \rightarrow (Ppxy \rightarrow \sim Np))$. a, AC.

c. $Cyy'\& \sim (x \subset y') \rightarrow (\exists a)(Paxy\& \sim Na)$. b, 3.

Hence theorem from c, AP.

9. $x \subset y\&y \subset x \rightarrow x = y$. AE, 2. 9, DE.

10. $(\exists x)(\exists y)(\sim (x = y))$. I(a), AE, 2. 9, DE.

To show that we have Boolean algebra we may now make comparison with the Huntington postulates previously mentioned.[4] For the nine [5] postulates essential in Huntington's formulation we have, among the theorems (or axioms) above, either exact analogues, or slightly simplified and stronger theorems from which exact analogues are immediately deducible. The correspondence is as follows:

Huntington Postulates	Present Theorems	Huntington Postulates	Present Theorems
1	1	7	3
2	9	8	4
3	II(b)	9	8
4	5	10	10
5	7		

[4] Huntington [2], pp. 288-309.

[5] His number 6, here omitted, he found redundant.

The converse deductions will be omitted, as it is well known that the theorems mentioned are valid in Boolean algebra.

PURDUE UNIVERSITY.

REFERENCES

1. L. Byrne, "Two brief formulations of Boolean algebra," *Bulletin of the American Mathematical Society*, vol. 52 (1946), pp. 269-272. Mentions other operational versions.

2. E. V. Huntington, "Sets of independent postulates for the algebra of logic," *Transactions of the American Mathematical Society*, vol. 5 (1904), pp. 288-309.

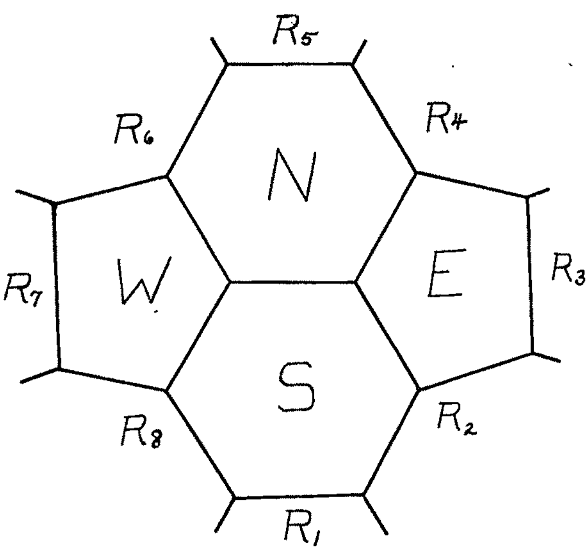
3. ———, "New sets of postulates for the algebra of logic," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 274-309, 557-558, 971. Includes extensive bibliography.

4. A. Tarski, "Zur Grundlegung der Boole'schen Algebra, I," *Fundamenta Mathematicae*, vol. 24 (1933), pp. 177-198.

ANOTHER REDUCIBLE EDGE CONFIGURATION.*

By Arthur Bernhart.

In any minimal five-color map all vertices are triple. Consider two adjacent regions N and S (primary components) forming vertices with E and W (guard components). Birkhoff (1) showed that such an edge configuration is reducible if all four components are pentagons. Franklin (2) showed the configuration is reducible if one primary component is a hexagon, the other three components pentagons. This paper proves the reducibility of two primary hexagons guarded by two pentagons.



The configuration forms the inside of an 8-ring $R_1R_2\cdots R_8$ and we choose the notation so that R_2 guards SE, and R_4 guards EN. Delete the edges R_2S, R_4N, R_6N and R_8S. The modified map must be four-colorable, and under the constraints $R_2 = R_8 \neq R_4 = R_6$. The outside of the original map must be colorable by at least one of the eighteen schemes which assigns one color to R_2 and R_8 and another color to R_4 and R_6. But since the whole map is not colorable, we may *exclude* from the outside any color scheme [on

* Received May 7, 1947.

the bounding ring] for which the known inside is colorable. This excludes twelve of the eighteen schemes outright, leaving six schemes for further study.

We indicate a color scheme by placing in brackets the colors assigned respectively to $R_1R_2 \cdots R_8$ and further abbreviate this notation by a number indicating the "alphabetical" order of the bracketed assignments in a complete list of the 274 possible 8-ring schemes. In this paper we shall refer only to these schemes:

23 [1212 3242]	111 [1231 2432]	156 [1232 1232]
59 [1213 2312]	115 [1231 3132]	158 [1232 1242]
64 [1213 2342]	117 [1231 3142]	172 [1232 1432]
95 [1231 2132]	137 [1231 4142]	237 [1234 2132]
97 [1231 2142]	150 [1231 4324]	248 [1234 2412]
104 [1231 2342]	151 [1231 4342]	253 [1234 2432]

A scheme marked X is *inside* colorable, one marked C is *outside* colorable, while N indicates *neither* side colorable. Contingent options obtainable by Kempe chain considerations will be indicated by separating the (and/or) alternatives by commas.

The outside must be colorable by at least one of these six schemes 64, 95, 97, 117, 137, and 248. But $64 \rightarrow (23X,\ 59X,\ 253X)$ therefore $64N$. By symmetry also $117N$ and $137N$ and $248N$. This leaves the outside option (95, 97). But $95 \rightarrow (97,\ 156X)$ and $97 \rightarrow (95,\ 158X)$ hence either alternative leads to both $95C$ and $97C$. However $104 \rightarrow (64N,\ 150X,\ 151X)$ hence $104N$. And $111 \rightarrow (104N,\ 172X)$ therefore $111N$, and by symmetry also $237N$. Finally $95 \rightarrow (111N,\ 115X,\ 237N)$ hence $95N$. But this is contrary to $95C$. The absurdity in coloring the outside proves that the inside configuration is reducible.

Using the term minor polygon to indicate pentagons and hexagons, we note that three edge configurations composed of minor polygons are now known to be reducible. An exhaustive investigation shows that other "minor" edge configurations with two or more pentagons are *not* reducible by this type of argument. For it is possible in each case to assign color schemes to the outside compatible with all known options. The reducible 8-ring considered here is the essential part in certain other reducible configurations listed by Winn (3). A further simplification is not possible however, since any three

of the four components, for instance ENS and ENW, form structures likewise not reducible.

UNIVERSITY OF OKLAHOMA,
NORMAN, OKLAHOMA.

BIBLIOGRAPHY

(1) G. D. Birkhoff, "The reducibility of maps," *American Journal of Mathematics*, vol. 35 (1912), pp. 115-128.

(2) Philip Franklin, "The four-color problem," *American Journal of Mathematics*, vol. 44 (1922), pp. 225-236.

(3) C. E. Winn, "On the minimum number of polygons in an irreducible map," *American Journal of Mathematics*, vol. 62 (1940), pp. 406-416.

AN APPLICATION OF ORTHONORMAL FUNCTIONS IN THE THEORY OF CONFORMAL MAPPING.*

By MENAHEM SCHIFFER.

Introduction. In this paper certain inequalities in the theory of conformal mapping of multiply-connected domains will be derived. The method of investigation is based on orthonormal functions and the kernel function which was introduced into the theory of functions by Bergman [1]. Although the main result of this paper is already known, it seems of interest to derive it in this new way in view of the great generality of the method of orthonormal functions, which opens an approach to similar results in the theory of harmonic functions of more than two variables and of more general differential equations of elliptic type. [Bergman-Schiffer].

1. Orthonormal systems. Let D be a domain in the complex z-plane which is bounded by n closed smooth curves C_ν $(\nu = 1, 2, \cdots n)$ and which contains the point at infinity. Consider the class Λ of all functions $f(z)$ which are regular in D, L^2-integrable over this domain, and possess a uniform integral $F(z)$ in D. It is well known that one may choose in an infinity of ways systems of functions $\{f_\nu(z)\}$ $(\nu = 1, 2, \cdots)$ of class Λ such that [1]

$$(1) \qquad \int\int_D f^*_\nu(z) f_\mu(z)\, dxdy = \delta_{\nu\mu}, \qquad \delta_{\nu\mu} = \begin{cases} 1 \text{ for } \nu = \mu \\ 0 \text{ for } \nu \neq \mu \end{cases}, \quad z = x + iy,$$

and that every function $f(z)$ of class Λ may be represented in a series

$$(2) \qquad f(z) = \sum_{\nu=1}^{\infty} A_\nu f_\nu(z)$$

which converges uniformly in every closed sub-domain of D. Such a system of functions $\{f_\nu(z)\}$ is called a complete orthonormal system with respect to the class Λ. We may construct the kernel function [Bergman 1]

$$(3) \qquad K(z, \zeta^*) = \sum_{\nu=1}^{\infty} f_\nu(z) f^*_\nu(\zeta)$$

which also converges uniformly in every closed sub-domain of D and which

* Received February 20, 1947.

[1] Throughout this paper we will denote the conjugate complex value of a by a^*.

is independent of the particular choice of the orthonormal system. The kernel function is closely related to certain mapping functions belonging to the domain D [Schiffer 2; cf. also Bergman 2].

As will be shown in **2**, it is possible to extend in a quite natural manner the definition of the class Λ so as to include several very important functions connected with canonical mappings of the domain D. We mention in particular the function

$$(4) \qquad \phi(z) = z + a_1 z^{-1} + a_2 z^{-2} + \cdots$$

which is regular in D with the exception of the pole at infinity, and maps D univalently upon a plane slit along segments parallel to the real axis. $\phi(z)$ has on every component C_ν of the boundary C of D constant imaginary parts, i. e.,

$$(4') \qquad \phi(z) = \alpha_\nu(z) + i\lambda_\nu \text{ for } z \varepsilon C_\nu, \alpha_\nu(z) \text{ and } \lambda_\nu \text{ real.}$$

Analogously, we define the function

$$(5) \qquad \psi(z) = z + b_1 z^{-1} + b_2 z^{-2} + \cdots$$

which maps D univalently upon a plane slit along segments parallel to the imaginary axis. Therefore one has now on every boundary continuum C_ν of D

$$(5') \qquad \psi(z) = \kappa_\nu + i\beta_\nu(z) \text{ for } z \varepsilon C_\nu,\ \kappa_\nu \text{ and } \beta_\nu(z) \text{ real.}$$

Now if we wish to include the derivatives $\phi'(z)$ and $\psi'(z)$ of these functions, which are of course uniform in D, into an orthonormal system, we encounter at once the difficulty that the integrals (1) do not converge. We shall therefore substitute instead of the orthonormalization condition (1) another which reduces to it in the case of functions of class Λ but remains defined for a much larger class of functions.

2. Generalization of the orthonormalization. We may transform the integral (1) over the domain D by integration by parts into an integral over the boundary C of D. We associate with every function $f_\nu(z)$ of the orthonormal system its integral function $F_\nu(z)$, i. e.,

$$(6) \qquad f_\nu(z) = (d/dz)F_\nu(z) = F'_\nu(z).$$

Then (1) may be written in the form

$$(7) \qquad (1/2i) \int_C F^*_\nu(z) F'_\mu(z)\, dz = \delta_{\nu\mu}$$

where the integral is taken in the positive sense with respect to D. Here we define

$$(7') \qquad (1/2i) \int_C F^*_\nu(z) F'_\mu(z)\, dz = \lim_{j\to\infty} (1/2i) \int_{\Gamma_j} F^*_\nu(z) F'_\mu(z)\, dz$$

where the Γ_j are a sequence of sets of smooth curves which converge to the set C in the Fréchet sense. The existence of the limit follows from the L^2-integrability of the $f_\nu(z)$ and from the identity

$$(7'') \qquad (1/2i) \int_C F^*_\nu(z) F'_\mu(z)\, dz = (1/2i) \int_{\Gamma_j} F^*_\nu(z) F'_\mu(z)\, dz + \int\int_{R_j} f^*_\nu(z) f_\mu(z)\, dxdy$$

where R_j is the ring domain enclosed by both sets of curves C and Γ_j.

Formulas (1) and (7) are equivalent if the $f_\nu(z)$ are of class Λ. But we may now introduce by means of (7) a metric into functional spaces of much greater generality than that formed by the family Λ.

Consider, for example, the space Ω of all functions which are regular in D, are L^2-integrable over every finite domain contained in D, and have uniform integral functions there. We do not assume that they are L^2-integrable over the entire domain D. We define for every pair of functions f and g in Ω the expressions (scalar product)

$$(8) \qquad (f^*, g) = (1/2i) \int_C F^*(z) G'(z)\, dz$$

where $F'(z) = f(z)$ and $G'(z) = g(z)$. In particular, we define the norm of f as

$$(9) \qquad N(f) = (f^*, f) = (1/2i) \int_C F^*(z) F'(z)\, dz.$$

For the sub-class Λ of Ω one derives from (1) that $N(f)$ is positive. In general, however, one can only assert that $N(f)$ is real. For integration by parts yields

$$(9') \qquad N^*(f) = -(1/2i)\int_C F(z)(F'(z))^*dz^* = (1/2i)\int_C F^*(z)F'(z)d$$
$$= N(f).$$

(In the same way, one proves immediately the symmetry property of the scalar product:

$$(9'') \qquad (f^*, g)^* = (g^*, f).)$$

We can, therefore, require as conditions of orthonormalization for a complete system $\{f_\nu(z)\}$ of class Ω only the equations

$$(10) \qquad (f^*_\nu, f_\mu) = \pm\delta_{\nu\mu}.$$

This is the generalization of the concept of orthonormal systems from the case of the class Λ to the class Ω which will be employed in this paper.

3. Construction of an orthonormal system for the class Ω. We start now to construct a complete orthonormal system for the functions of class Ω, i. e., a system of functions of class Ω orthonormalized in the sense of **2**, such that every function of class Ω may be developed into a series of these functions which converges uniformly in each closed subdomain of D. For this purpose, we have to investigate first the expressions (ϕ'^*, f) and (ψ'^*, f), where f is an arbitrary function of class Ω and ϕ, ψ are defined by (4) and (5) respectively. We find in view of (4'), if f is continuous on C,

$$(11) \qquad (\phi'^*, f) = (1/2i)\int_C \phi^*(z)f(z)\,dz = (1/2i)\sum_{\nu=1}^{n}\int_{C_\nu}(\alpha_\nu(z) - i\lambda_\nu)f(z)dz$$
$$= (1/2i)\sum_{\nu=1}^{n}\int_{C_\nu}(\alpha_\nu(z) + i\lambda_\nu)f(z)\,dz = (1/2i)\int_C \phi(z)f(z)\,dz.$$

The last integral may be evaluated by means of the residue theorem. Assuming for $f(z)$ a development at infinity of the form

$$(12) \qquad f(z) = c_0 + c_2z^{-2} + c_3z^{-3} + \cdots$$

we get in view of (4)

$$(13) \qquad (\phi'^*, f) = -\pi(a_1c_0 + c_2).$$

The same result holds for the most general function f of class Ω. For every function of class Ω may be approximated by functions of this class which are continuous on the boundary C. The approximation is uniform in every closed subdomain of D and is in the mean valid for the whole domain D [Farrell]. Hence, we may conclude from (7'') the validity of (13) for

general functions of class Ω, once it has been proved for functions of the special type.

Analogously, we obtain from (5′), for f continuous on C,

$$(14)\quad \begin{aligned}(\psi'^*, f) &= (1/2i) \int_C \psi^*(z) f(z)\, dz = (1/2i) \sum_{\nu=1}^{n} \int_{C_\nu} (\kappa_\nu - i\beta_\nu(z)) f(z)\, dz \\ &= -(1/2i) \sum_{\nu=1}^{n} \int_{C_\nu} (\kappa_\nu + i\beta_\nu(z)) f(z)\, dz = -(1/2i) \int_C \psi(z) f(z)\, dz.\end{aligned}$$

Hence the residue theorem yields in view of (5) and (12)

$$(15)\qquad (\psi'^*, f) = \pi(b_1 c_0 + c_2).$$

Again we extend the validity of (15) to all functions f of class Ω by exactly the same argument as before.

Consider now the function

$$(16)\qquad U(z) = \tfrac{1}{2}(\phi(z) + \psi(z)) = z + \tfrac{1}{2}(a_1 + b_1) z^{-1} + \cdots$$

the derivative of which is of class Ω. In view of (13) and (15) we have for every $f \varepsilon \Omega$

$$(17)\qquad (U'^*, f) = -\tfrac{1}{2}\pi(a_1 - b_1) c_0.$$

The factor

$$(18)\qquad S(D) = a_1 - b_1$$

appearing in (17) plays an important role in the theory of conformal mapping. It has been called the span of the domain D [Schiffer, 1], is always non-negative and a conformal invariant with respect to mappings $w(z)$ which are normalized at infinity as

$$(19)\qquad w(z) = z + k_0 + k_1 z^{-1} + \cdots.$$

If $f(z)$ is, in particular, of class Λ, we have $c_0 = 0$ and, therefore:

THEOREM I. *The function $U'(z) = \frac{1}{2}(\phi'(z) + \psi'(z))$ is orthogonal to every function of class Λ and only to functions of class Ω which belong to the class Λ.*

We remark further that $U'(z)$ has a negative norm, namely, in view of (17)

$$(20)\qquad N(U') = -\tfrac{1}{2}\pi S(D).$$

Let us consider next the function

$$V(z)=\tfrac{1}{2}(\phi(z)-\psi(z))=\tfrac{1}{2}S(D)z^{-1}+\cdots \tag{21}$$

the derivative $V'(z)$ of which is certainly of class Λ. In view of (13) and (15), we find

$$(V'^{*}(z),f(z))=-\tfrac{1}{2}\pi((a_1+b_1)c_0+2c_2). \tag{22}$$

Considering only functions f of class Λ, we have $c_0=0$ and hence:

THEOREM II. *The function* $V'(z)=\frac{1}{2}(\phi'(z)-\psi'(z))$ *is orthogonal to every function of class* Λ *which vanishes at infinity at least of third order and only to those functions of class* Λ.

The norm of $V'(z)$ is certainly positive, since $V'(z)$ is of class Λ. In fact, we have from (21) and (22)

$$(V'^{*},V')=\tfrac{1}{2}\pi S(D). \tag{23}$$

This equation gives, by the way, a new proof of the fact that $S(D)$ is a non-negative number.

Now we construct easily the following complete orthonormal system with respect to Ω:

$$f_1(z)=(2/\pi S)^{\frac{1}{2}}U'(z),\ f_2(z)=(2/\pi S)^{\frac{1}{2}}V'(z),\ f_3(z),\ f_4(z),\cdots \tag{24}$$

where $f_3,f_4,\cdots$ vanish at infinity at least of the third order in view of Theorems I and II, and where $f_2,f_3,\cdots$ are a complete orthonormal system with respect to the class Λ.

This system will be used in several applications to the theory of conformal mapping.

4. The kernel function of an orthonormal system in Ω. We associate with a given complete system $g_1(z),g_2(z)\cdots$ of functions of class Ω which is orthonormalized by conditions (10), the kernel function

$$k(z,\zeta^{*})=\sum_{\nu=1}^{\infty}\pm g_\nu(z)g^{*}{}_\nu(\zeta) \tag{25}$$

where a plus sign stands before functions g_ν with positive norm, and a minus sign before functions g_ν with negative norm. In the particular case (24) we have

$$k(z,\zeta^{*})=-(2/\pi S)U'(z)U'^{*}(\zeta)+(2/\pi S)V'(z)V'^{*}(\zeta)+\sum_{\nu=3}^{\infty}f_\nu(z)f^{*}{}_\nu(\zeta \tag{26}$$

We prove now the following:

THEOREM III. *In every orthonormal system $\{g_\nu(z)\}$ of class Ω there appears exactly one function, say $g_1(z)$, with negative norm.*

Proof. Suppose there were two functions $g_1(z)$ and $g_2(z)$ with negative norm. We could find two complex numbers λ_1 and λ_2, neither of them zero, such that the linear combination

$$\gamma(z) = \lambda_1 g_1(z) + \lambda_2 g_2(z) \tag{27}$$

vanishes at infinity at least of second order and is therefore of class Λ. Thus the norm of $\gamma(z)$ is positive, while from (27) we derive

$$N(\gamma) = |\lambda_1|^2 N(g_1) + |\lambda_2|^2 N(g_2) < 0, \tag{28}$$

which leads to a contradiction.

On the other hand, there must exist at least one $g_\nu(z)$ with negative norm, since $U'(z)$ may be composed of $g_\nu(z)$ and has a negative norm. This completes the proof.

We may characterize the kernel function by the following two properties:

a) $$k(z, \zeta^*) = k^*(\zeta, z^*).$$

b) If $f(z)$ is an arbitrary function of class Ω, we have the identity

$$f(z) = (k^*(\zeta, z^*), f(\zeta)). \tag{29}$$

The last result follows at once from the definition (25) and the conditions of orthonormalization (10).

If there were another function $L(z, \zeta^*)$ of class Ω satisfying both these conditions, one would have, by applying (29), first with respect to k and then to L

$$\begin{aligned} L(z, \zeta^*) &= (k^*(t, z^*), L(t, \zeta^*)) = (L^*(t, \zeta^*), k(t, z^*)) \\ &= k^*(\zeta, z^*) = k(z, \zeta^*). \end{aligned} \tag{30}$$

This shows that the conditions a) and b) determine the kernel function in a unique manner and that, therefore, it is the same function independently of the orthonormal system from which it is constructed.

We shall now use this fact in order to construct the kernel function by means of two different orthonormal systems and obtain interesting results by comparing the two different representations so obtained. As one system we use the set (24) leading to the representation (26). As the other system,

we use an arbitrary complete system for the class Λ to which we adjoin the function 1, of class Ω. We may orthonormalize this system and get in this way a complete orthonormal system for the class Ω. We remark at first that the norm of 1 is easily computed:

$$N(1) = (1/2i) \int_C z^* dz = -E(D) \tag{31}$$

where $E(D)$ is the finite area enclosed by the set of smooth curves C_ν, i.e., $E(D)$ is the area of the complementary part to D of the plane.

Thus, we get a system

$$g_1(z) = E(D)^{-\frac{1}{2}}, g_2(z), g_3(z), \cdots \tag{32}$$

with corresponding kernel function

$$k(z, \zeta^*) = -E(D)^{-1} + \sum_{\nu=2}^{\infty} g_\nu(z) g^*_\nu(\zeta) \tag{33}$$

where all the signs in the right-hand sum are positive in view of Theorem III.

In particular, we obtain from (26) and (33)

$$k(\infty, \infty^*) = -(2/\pi S) = -E(D)^{-1} + \sum_{\nu=2}^{\infty} | g_\nu(\infty) |^2 \tag{34}$$

since all the functions $f_\nu(z)$ of class Λ in (24) vanish at infinity.

Thus we are led to the inequality

$$S(D) \geq (2/\pi) E(D). \tag{35}$$

This inequality has been derived formerly by variational methods [Schiffer 1; cf. also Grunsky, pp. 139-140]. It appears of very great interest that this result comes out as a by-product of our investigations of orthonormal functions. It is obvious that the estimate may be improved by taking into account further $g_\nu(z)$ and so the result may be extended.

On the other hand, it is not easy to show by our method that (35) is the best possible inequality. In order to show this fact, one has to study the possibility of equality in (35). One sees from our considerations above that equality can hold only if the functions $g_2(z), g_3(z) \cdots$ in (32) all vanish at infinity. They represent in this case a complete orthonormal system for the class Λ and the constant $g_1(z)$ is orthogonal to all functions of Λ. Thus one has necessarily

$$g_1(z) = E(D)^{-\frac{1}{2}} = (2/\pi S)^{\frac{1}{2}} U'(z) = f_1(z) \tag{36}$$

i. e., in view of the equality in (35)

$$U'(z) = 1, \qquad U(z) = \tfrac{1}{2}(\phi(z) + \psi(z)) = z. \tag{36'}$$

By variational methods it was possible to show that $U(z)$ is for every given domain D a univalent function and maps it, therefore, upon a domain for which (36′) holds. This result is much more difficult to obtain by our method. On the other hand, the theory of orthonormal functions is easily extended to more general differential equations and to a greater number of variables, so that every result obtained by means of it is of great interest from the methodical viewpoint.

5. The kernel function of an orthonormal system in Λ. Another interesting application is the following: Take the function $V'(z)$ defined in (21) which is of class Λ, and consider a complete orthonormal system $\{h_\nu(z)\}$ with respect to Λ, containing $h_1(z) = (2/\pi S)^{\frac{1}{2}} V'(z)$ as first element (cf. (23))

$$h_1(z) = (2/\pi S)^{\frac{1}{2}} V'(z), h_2(z), h_3(z), \cdots. \tag{37}$$

In view of Theorem II, all the functions $h_\nu(z)$ $(\nu = 2, 3, \cdots)$ vanish at infinity at least of the third order.

If we construct, therefore, the kernel function $K(z, \zeta^*)$ corresponding to the class Λ by means of the system (37) we get

$$K(z, \zeta^*) = (2/\pi S)\, V'(z) V'^*(\zeta) + \sum_{\nu=2}^{\infty} h_\nu(z) h^*{}_\nu(\zeta) \tag{38}$$

and, in view of (21) and the behavior of the $h_\nu(z)$ $(\nu = 2, 3, \cdots)$ at infinity,

$$\lim_{\zeta\to\infty} \zeta^{*2} K(z, \zeta^*) = -(1/\pi) V'(z). \tag{39}$$

On the other hand, we derive from (26) and (16)

$$k(z, \infty^*) = -(2/\pi S)\, U'(z). \tag{40}$$

Thus we get finally from (16), (21), (39) and (40),

$$S(D) = -(2/\pi)[k(\infty, \infty^*)]^{-1}, \tag{41}$$

$$\phi'(z) = -\pi \lim_{\zeta\to\infty} [\tfrac{1}{2} S(D) k(z, \zeta^*) + \zeta^{*2} K(z, \zeta^*)], \tag{41'}$$

$$\psi'(z) = -\pi \lim_{\zeta\to\infty} [\tfrac{1}{2} S(D) k(z, \zeta^*) - \zeta^{*2} K(z, \zeta^*)]. \tag{41''}$$

Since the calculation of the kernel function is a known algorithm, we have obtained by (41), (41′) and (41″) a new method for mapping a given domain D upon slit domains of canonical type.

HARVARD UNIVERSITY.

BIBLIOGRAPHY

S. Bergman: 1. *Partial differential equations (Conformal mapping of multiply-connected domains)*, Publication of Brown University, Providence, R. I., 1941.

——— 2. "A remark on the mapping of multiply-connected domains," *American Journal of Mathematics*, vol. 68 (1946), pp. 20-28.

S. Bergman-M. Schiffer: "A representation of Green's and Neumann's functions in the theory of differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 609-638.

O. J. Farrell: "On approximation to an analytic function by polynomials," *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 908-914.

H. Grunsky: "Neue Abschätzungen zur konformen Abbildung ein- und mehrfach zusammenhängender Gebiete," *Schriften des mathematischen Seminars Berlin*, vol. 1 (1932), pp. 95-140.

M. Schiffer: 1. "The span of multiply-connected domains," *Duke Mathematical Journal*, vol. 10 (1943), pp. 209-216.

——— 2. "The kernel function of an orthonormal system," *Duke Mathematical Journal*, vol. 13 (1946), pp. 529-540.

SOME LIMIT THEOREMS.*[1]

By C. T. RAJAGOPAL.

In this note I record relations between certain classical limit theorems which I believe have not been noticed before. My main theorems (3) make use of the following

NOTATION. $\{s_n\}$ is any sequence real or complex, unless it is explicitly stated to be real. $a_n > 0$, $d_n > 0$ $(n = 1, 2, \cdots)$ are such that

$$\sum_{\nu=1}^{n} a_\nu = A_n \to \infty, \qquad \sum_{\nu=1}^{n} d_\nu = D_n \to \infty \qquad (n \to \infty).$$

$$\sigma(s_n, a_n) = \frac{\sum_{\nu=1}^{n} a_\nu s_\nu}{A_n}, \qquad \sigma(s_n, d_n) = \frac{\sum_{\nu=1}^{n} d_\nu s_\nu}{D_n}.$$

Some well-known theorems of Cesàro, Hardy [2] and Rademacher [6] follow from the relative asymptotic distribution of the two sequences of points $\sigma(s_n, a_n)$, $\sigma(s_n, d_n)$ in special cases.

1. I begin by proving a lemma which is equivalent to one of my main theorems (Theorem 3).

LEMMA 1. *Let $\{U_n\}$ be any real sequence. Let $\{M_n\}$ be a positive monotonic sequence and $\{\Lambda_n\}$ an increasing divergent sequence. Let*

$$\bar{\underline{\lambda}} = \overline{\underline{\lim}}_{n\to\infty} \{ \sum_{1}^{n} (U_\nu - U_{\nu-1}) M_\nu / \sum_{1}^{n} (\Lambda_\nu - \Lambda_{\nu-1}) M_\nu \}.^2$$

Then (i) *in case $\{M_\nu\}$ is increasing, $\overline{\underline{\lim}}$ (U_n/Λ_n) lie in the closed interval* $(\underline{\lambda}, \bar{\lambda})$.

(ii) *In case $\{M_\nu\}$ is decreasing and subject to the condition*

(C) $$\{\sum_{1}^{n} (\Lambda_\nu - \Lambda_{\nu-1}) M_\nu\}/\Lambda_n < k M_n{}^{3} \qquad (k = a\ constant),$$

* Received April 1, 1947.

[1] I am indebted to Dr. V. Ganapathy Iyer and Mr. T. K. Raghavachari for helping me to clear up several points.

[2] Any number with suffix 0 is taken to be 0.

[3] $k > 1$ since $\sum_{1}^{n} (\Lambda_\nu - \Lambda_{\nu-1}) M_\nu < k \sum_{1}^{n} (\Lambda_\nu - \Lambda_{\nu-1}) M_n \leq k \sum_{1}^{n} (\Lambda_\nu - \Lambda_{\nu-1}) M_\nu$.

$\overline{\lim}\ (U_n/\Lambda_n)$ *lie in the closed interval whose end-points divide the segment* $(\underline{\lambda}, \bar{\lambda})$ *in the ratios* $-k'$, $-1/k'$, *where* k' *is such that* $1/k + 1/k' = 1$.

Proof. Let us write $V_n = \sum_1^n (U_\nu - U_{\nu-1})M_\nu$, $D_n = \sum_1^n (\Lambda_\nu - \Lambda_{\nu-1})M_\nu$. Then, given $\epsilon > 0$, we can find N so that $\lambda - \epsilon < V_n/D_n < \bar{\lambda} + \epsilon$ for $n \geq N$. Also by partial summation

$$U_n = \sum_1^n \frac{V_\nu - V_{\nu-1}}{M_\nu} = \sum_1^{N-1} V_\nu \left(\frac{1}{M_\nu} - \frac{1}{M_{\nu+1}}\right) + \sum_N^{n-1} \cdots + \frac{V_n}{M_n}.$$

In case (i)

$$U_n > \sum_1^{N-1} V_\nu \left(\frac{1}{M_\nu} - \frac{1}{M_{\nu+1}}\right) + (\underline{\lambda} - \epsilon)\left[\sum_N^{n-1} D_\nu \left(\frac{1}{M_\nu} - \frac{1}{M_{\nu+1}}\right) + \frac{D_n}{M_n}\right]$$

$$= \sum_1^{N-1} \cdots + (\underline{\lambda} - \epsilon)\left[\sum_{n+1}^{n} \frac{D_\nu - D_{\nu-1}}{M_\nu} + \frac{D_n}{M_n}\right]$$

or

$$\frac{U_n}{\sum_1^n (D_\nu - D_{\nu-1})/M_\nu} > \frac{k_1}{\sum_1^n (D_\nu - D_{\nu-1})/M_\nu} + \underline{\lambda} - \epsilon \qquad (k_1 = \text{a constant})$$

$$> \underline{\lambda} - 2\epsilon$$

for all sufficiently large n, in virtue of the divergence of $\sum (D_n - D_{n-1})/M_n \equiv \sum(\Lambda_n - \Lambda_{n-1})$. ϵ being arbitrary, this means $\underline{\lim}\ (U_n/\Lambda_n) \geq \underline{\lambda}$. The proof can now be completed by a repetition of the preceding argument involving $\bar{\lambda}$ instead of $\underline{\lambda}$.

In case (ii)

$$U_n > \sum_1^{N-1} V_\nu \left(\frac{1}{M_\nu} - \frac{1}{M_{\nu+1}}\right) + (\bar{\lambda} + \epsilon)\left[\sum_N^{n-1} D_\nu \left(\frac{1}{M_\nu} - \frac{1}{M_{\nu+1}}\right)\right] + (\underline{\lambda} - \epsilon)\frac{D_n}{M_n}$$

$$= \sum_1^{N-1} \cdots + (\bar{\lambda} + \epsilon)\left[\frac{D_n}{M_n} + \sum_{N+1}^{n} \frac{D_\nu - D_{\nu-1}}{M_\nu} - \frac{D_n}{M_n}\right] + (\underline{\lambda} - \epsilon)\frac{D_n}{M_n}$$

or

$$\frac{U_n}{\sum_1^n (D_\nu - D_{\nu-1})/M_\nu} > \frac{k_2}{\sum_1^n (D_\nu - D_{\nu-1})/M_\nu} + (\underline{\lambda} - \bar{\lambda} - 2\epsilon)\frac{D_n/M_n}{\sum_1^n (D_\nu - D_{\nu-1})/M_\nu}$$

$$+ \bar{\lambda} + \epsilon \qquad (k_2 = \text{a constant}).$$

Now (C) is equivalent to $\dfrac{D_n/M_n}{\sum_1^n (D_\nu - D_{\nu-1})/M_\nu} < k$ and therefore

$$\frac{U_n}{\Lambda_n} > -\epsilon + (\underline{\lambda} - \bar{\lambda} - 2\epsilon)k + \bar{\lambda} + \epsilon$$

for all sufficiently large n, whence

$$\underline{\lim}\,\frac{U_n}{\Lambda_n} \geq (\underline{\lambda}-\bar{\lambda})k+\bar{\lambda}.$$

By establishing in a similar manner

$$\underline{\lim}\,\frac{U_n}{\Lambda_n} \leq (\bar{\lambda}-\underline{\lambda})k+\lambda,$$

we can complete the proof.

Remarks. (1) $\sum(\Lambda_n-\Lambda_{n-1})\mathrm{M}_n \equiv \sum(D_n-D_{n-1})$ *is a divergent series of positive terms whether* $\{\mathrm{M}_n\}$ *is increasing or decreasing.* For, in the former case the result is trivial; while, in the latter case, the convergence of $\sum(\Lambda_n-\Lambda_{n-1})\mathrm{M}_n$, if possible, will imply that of

$$\sum(\Lambda_n-\Lambda_{n-1})\mathrm{M}_n/\{\sum_{\nu=1}^{n}(\Lambda_\nu-\Lambda_{\nu-1})\mathrm{M}_\nu\}$$

and hence, by (C), that of $\sum(\Lambda_n-\Lambda_{n-1})/\Lambda_n$ which is impossible.

(2) *When* $\{\mathrm{M}_n\}$ *is increasing, we have*

$$\underline{\lim}\,\frac{U_n}{\Lambda_n}\begin{cases} \leq \overline{\lim}\,\dfrac{\sum_1^n (U_\nu-U_{\nu-1})\mathrm{M}_\nu}{\sum_1^n(\Lambda_\nu-\Lambda_{\nu-1})\mathrm{M}_\nu} \leq \overline{\lim}\,\dfrac{U_n-U_{n-1}}{\Lambda_n-\Lambda_{n-1}} \\ \geq \underline{\lim} \qquad\quad `` \qquad\quad \geq \underline{\lim} \qquad `` \end{cases}$$

which is an augmentation of the inequalities giving rise to the Cauchy-Stolz limit theorem, obtained by using these very inequalities in conjunction with Lemma 1.

2. My first theorem takes $\mathrm{M}_n=D_n$ in Lemma 1 and obtains a necessary and sufficient condition for the finite oscillation of U_n, extending a result of Kronecker and Knopp.

THEOREM 1. *Let* $\sum u_n$ *be a real series and* $\{D_n\}$ *a positive increasing divergent sequence. Let*

$$U_n=\sum_1^n u_\nu,\ V_n=\sum_1^n u_\nu D_\nu,\ \underline{\bar{U}}=\underline{\overline{\lim}}_{n\to\infty}\,U_n,\ \underline{\bar{\lambda}}=\underline{\overline{\lim}}_{n\to\infty}\,(V_n/D_n).$$

(a) *If* $\underline{\bar{U}}$ *are finite, then*

$$-(\bar{U}-\underline{U})\leq\underline{\lambda}\leq 0\leq\bar{\lambda}\leq(\bar{U}-\underline{U}).$$

(b) *Conversely, if* $|\lambda| \leq K$, *a constant independent of* $\{D_n\}$, *then* $\bar{U}$ *are finite and* $\bar{U} - \underline{U} \leq 2K$.

Proof (a) Taking $M_n = D_n$ in Lemma 1, we first note that $\underline{\lambda} \leq 0 \leq \bar{\lambda}$.

Next we prove $\bar{\lambda} \leq \bar{U} - \underline{U}$ supposing that $\bar{\lambda} > 0$ (the case $\bar{\lambda} = 0$ being trivial). Given an arbitrary positive $\epsilon < \bar{\lambda}/2$ we can find m and $m' \geq m+1$ so that

$$| U_{m'} - U_m | < \bar{U} - \underline{U} + 2\epsilon$$

and

$$\frac{V_m}{D_m} < \epsilon \leq \frac{\max}{\min}\left(\frac{V_{m+1}}{D_{m+1}}, \frac{V_{m+2}}{D_{m+2}}, \cdots, \frac{V_{m'-1}}{D_{m'-1}}\right) \leqq \bar{\lambda} - \epsilon < \frac{V_{m'}}{D_{m'}}$$

or, since $V_\nu - V_m > 0$ $(m+1 \leq \nu \leq m')$,[4]

$$\sum_{m+1}^{m'-1} (V_\nu - V_m)\left(\frac{1}{D_\nu} - \frac{1}{D_{\nu+1}}\right) + (V_{m'} - V_m)\frac{1}{D_{m'}} > (V_{m'} - V_m)\frac{1}{D_{m'}} > \bar{\lambda} - 2\epsilon$$

Since the extreme left-hand member of the last relation is identically equal to $U_{m'} - U_m$, the relation gives

$$\bar{U} - \underline{U} + 2\epsilon > \bar{\lambda} - 2\epsilon$$

i. e.,

$$\bar{U} - \underline{U} \geq \bar{\lambda}.$$

The proof is now completed by changing u_n to $-u_n$ and so obtaining $\bar{U} - \underline{U} \geq -\bar{\lambda}$.

(b) From the hypothesis and the identity

$$U_{n'} - U_n = -\frac{V_n}{D_{n+1}} + \sum_{n+1}^{n'-1} V_\nu\left(\frac{1}{D_\nu} - \frac{1}{D_{\nu+1}}\right) + \frac{V_{n'}}{D_{n'}}, \quad n' \geqq n+1,$$

we know there is an N such that, for $n \geq N$,

$$| U_{n'} - U_n | < 2(K+\epsilon) + \sum_{n+1}^{n'-1} (K+\epsilon) D_\nu \left(\frac{1}{D_\nu} - \frac{1}{D_{\nu+1}}\right)$$

$$< 2(K+\epsilon) + (K+\epsilon)\log(D_{n'}/D_{n+1}).$$

Let n, n' tend to infinity through the arbitrary sequences of values p_r, p'_r $(r \to \infty)$ respectively:

$$p_1 < p'_1 < p_2 < \cdots < p_r < p'_r < p_{r+1} < \cdots$$

---

[4] This is obvious when $V_m \leqq 0$. When $V_m > 0$, $V_\nu/D_\nu > V_m/D_m$ and so $V_\nu > V_m D_\nu/D_m > V_m$.

and choose

$$D_\nu = r + \frac{\nu - p_r}{2(p'_r - p_r)}, \qquad p_r < \nu \leq p'_r,$$

$$D_\nu = r + \tfrac{1}{2} + \frac{\nu - p'_r}{2(p_{r+1} - p'_r)}, \qquad p'_r < \nu \leq p_{r+1},$$

so that $D_{p'_r}/D_{p_{r+1}} \to 1$ as $r \to \infty$. Then

$$| U_{n'} - U_n | < 2(K + \epsilon) + (K + \epsilon)\epsilon, \qquad n' > n \to \infty,$$

whence the desired conclusion follows at once.

COROLLARY. *A necessary and sufficient condition for the convergence of any series* $\sum u_n$, *real or complex, is* $V_n/D_n = o(1)$, $n \to \infty$, *for every positive increasing divergent sequence* $\{D_n\}$. [Kronecker-Knopp, **4**, Satz 1.]

As a complement to this result we have the next theorem in which $\{D_n\}$ is a particular sequence.

THEOREM 2. *In the notation of Theorem* 1, *necessary and sufficient conditions for the convergence of* $\sum u_n$ *are* $V_n/D_n = o(1)$ *and the convergence of* $\sum V_n(D_n^{-1} - D_{n+1}^{-1})$.

The proof is obvious from the identity on which the proof of Theorem 1(b) is based.

Now $\sum(D_n^{-1} - D_{n+1}^{-1})f(D_n^{-1})$ is convergent when $f(x) \uparrow \infty$ as $x \downarrow 0$ and $\int_0 f(x)dx$ is convergent [**8**, p. 399, ex. 5]. Hence we can take

$$V_n = O\{f(D_n^{-1})\}, \qquad f(x) = x^{-1}g(x^{-1})$$

and obtain the

COROLLARY. $V_n/D_n = O\{g(D_n)\}$ *is sufficient for the convergence of* $\sum u_n$ *when* $g(x)$ *satisfies the conditions*: $g(x) \to +0$ *and* $xg(x) \uparrow \infty$ *as* $x \uparrow \infty$. $\int^\infty g(x)dx/x$ *is convergent.* [Karamata, 5, 2°.]

In particular we can have either $g(x) = x^{-\rho}$, $\rho > 0$, *or*

$$g(x) = 1/L_k(x)(\log_k x)^\rho, \qquad \rho > 0.$$

where $L_k(x) = \log x \log_2 x \cdots \log_k x$. [The case $g(x) = x^{-\rho}$, $\rho \geq 1$, is trivial and not a part of the result stated above.]

3. THEOREM 3. *In the notation explained at the outset, let* $\sigma(s_n, a_n)$, $\sigma(s_n, d_n)$ *be defined for a real sequence* $\{s_n\}$.

(i) *If a_n/d_n is decreasing, then* $\overline{\lim}\,\sigma(s_n, a_n)$ *lie in the closed interval with end-points* $\overline{\underline{\lim}}\,\sigma(s_n, d_n)$.

(ii) *If a_n/d_n is increasing and subject to the condition $a_n/d_n < kA_n/D_n$, then* $\overline{\underline{\lim}}\,\sigma(s_n, a_n)$ *lie in the closed interval whose end-points divide the segment between* $\overline{\underline{\lim}}\,\sigma(s_n, d_n)$ *in the ratios $-k'$, $-1/k'$, k' being defined by* $1/k + 1/k' = 1$.

Proof. In Lemma 1, put $\Lambda_n - \Lambda_{n-1} = a_n$, $(\Lambda_n - \Lambda_{n-1})M_n = d_n$, $U_n - U_{n-1} = a_n s_n$ so that $(U_n - U_{n-1})M_n = d_n s_n$.

COROLLARY 1. *If a_n and d_n are as in either* (i) *or* (ii) *of Theorem* 3, *then* $\lim \sigma(s_n, d_n) = 0$ *implies* $\lim \sigma(s_n, a_n) = 0$. [Cesàro-Hardy, **2**, § 4.]

COROLLARY 2. *Suppose, in Theorem* 3, $\lim \sigma(s_n, a_n) = 0$. *Then in case* (i) *of the theorem,* $\overline{\underline{\lim}}\,\sigma(s_n, d_n)$ *are not of the same sign i. e.,*

$$\underline{\lim}\,\sigma(s_n, d_n) \leq 0 \leq \overline{\lim}\,\sigma(s_n, d_n)\,;$$

while, in case (ii), *either* $\overline{\underline{\lim}}\,\sigma(s_n, d_n)$ *are not of the same sign, or (if they are) the ratio of the numerically smaller to the larger does not exceed* $1/k'$.[5]

Rademacher has proved a theorem [**6**, Satz I] which is essentially the case (i) of the above corollary for $d_n \equiv 1$. A result supplementary to this case is the following.

THEOREM 4. *In the notation of Theorem* 3, *on the hypothesis $a_n \mid d_n \downarrow 0$, a sufficient condition for $\sigma(s_n, d_n)$ to tend to* 0, *whenever $\{s_n\}$ makes $\sum a_n s_n$ convergent, is* $\lim\,(a_n D_n \mid d_n) > 0$.

When d_n is bounded, the condition is also necessary.

Proof. The sufficiency part of the theorem follows from Lemma 2 below due to Rademacher [**6**, Satz III], and the necessity part from Lemma 3 due to Fuchs in the case $d_n \equiv 1$ [**1**, Theorem 1].

Note. One can show, by examples, that even when d_n is unbounded, the condition $a_n/d_n \downarrow 0$, without $\underline{\lim}\,(a_n D_n/d_n) > 0$. cannot always ensure $\lim \sigma(s_n, d_n) = 0$ for every sequence $\{s_n\}$ which makes $\sum a_n s_n$ convergent.

[5] To prove the second alternative of case (ii), suppose first that $\overline{\underline{\lim}}\,\sigma(s_n, d_n) = \overline{\underline{\lambda}}$ are both positive. Then the last two inequalities in the proof of Lemma 1 reduce to

$$0 \geq (\underline{\lambda} - \bar{\lambda})k + \bar{\lambda}, \qquad 0 \leq (\bar{\lambda} - \underline{\lambda})k + \underline{\lambda}.$$

Of the above inequalities, the second gives $\bar{\lambda}/\underline{\lambda} \geq (k-1)/k$ which is trivial; while the first gives $\bar{\lambda}/\underline{\lambda} \geq k/(k-1) = k'$ which is the result stated. The case in which $\overline{\underline{\lambda}}$ are both negative is dealt with in the same way.

For instance, when $a_n = 1/(n+1)\log(n+1)$, $d_n = n$, $\{s_n\}$ is $2, \frac{1}{3}, -4, 5, \frac{1}{6}, -7, \cdots$, $\sum a_n s_n$ is convergent but $\underline{\lim}\,\sigma(s_n, d_n) = -\frac{4}{3}$ and $\overline{\lim}\,\sigma(s_n, d_n) = \frac{2}{3}$; although with the same a_n and d_n, $\sum a_n s'_n$ is convergent and $\lim \sigma(s'_n, d_n) = 0$ if $ns'_n = s_n$ defined as above.

LEMMA 2. *If $a_n/d_n \downarrow 0$, then (without any other restriction on a_n and d_n) the convergence of $\sum a_n s_n$ involves $\sum_1^n d_\nu s_\nu = o(d_n/a_n)$.*

LEMMA 3. *If $d_n \leq \delta$, $a_n/d_n \downarrow 0$, $\underline{\lim}\,(a_n D_n/d_n) = 0$, we can find a real sequence $\{s_n\}$ for which $\sum a_n s_n$ is convergent and $\overline{\lim}\,\sigma(s_n, d_n) > 0$.*

Proof. Setting $a_n D_n/d_n = b(n)$, we can find a null sequence $\{b(n_r)\}$ such that

$$\sum_{r=1}^{\infty} b(n_r) < \infty$$

and

$$[D_{n_1}] > 1, \qquad n_{r+1} \geq n_r + [D_{n_r}] \equiv N_r \qquad (r = 1, 2, 3, \cdots)$$

where $[D_{n_r}]$ denotes the integral part of D_{n_r}.

Choose $s_\nu = 1/d_\nu$ if ν is in one of the intervals $n_r \leq \nu < N_r$; $s_\nu = (-1)^\nu/d_\nu$ otherwise. Then the partial remainder $\sum_m^N a_\nu s_\nu$ will consist of numerically decreasing terms, with groups of terms alternating in sign separated by groups of terms all positive. If $N_{k-1} \leq m < n_k$, the contribution of the former groups is numerically less than

$$a_m/d_m + \sum_{r\geq k} a_{N_r}/d_{N_r} < a_m/d_m + \sum_{r\geq k} a_{n_r}/d_{n_r} < a_m/d_m + \sum_{r\geq k} b(n_r).$$

Also, the contribution of one of the groups of positive terms is

$$\sum_{n_r \leq \nu < N_r} a_\nu/d_\nu < [D_{n_r}] a_{n_r}/d_{n_r} \leq b(n_r).$$

Hence

$$\sum_{\nu=m}^{N} a_n s_n = O\{a_m/d_m + \sum_{r\geq k} b(n_r)\} = o(1), \qquad m \to \infty.$$

Thus $\sum a_n s_n$ is convergent. But

$$\sigma(s_{N_r}, d_{N_r}) \geq \frac{([D_{n_1}] - 1) + \cdots + ([D_{n_r}] - 1)}{D_{N_r}}$$

which gives, in virtue of the condition $D_{N_r} - D_{n_r} \leq \delta(N_r - n_r) = \delta[D_{n_r}]$,

$$\sigma(s_{N_r}, d_{N_r}) > \frac{[D_{n_r}] - 1}{D_{n_r} + \delta[D_{n_r}]}$$

whence

$$\overline{\lim}\, \sigma(s_n, d_n) \geq \frac{1}{1+\delta}.$$

THEOREM 5. *If $a_n/d_n \downarrow 0$, then a sufficient condition for the convergence of $\sum a_n s_n$ is $\sum_1^n d_\nu s_\nu = O\{G(d_n/a_n)\}$, $n \to \infty$, where $G(x)$ is such that $G(x)/x \to +0$ and $G(x) \uparrow \infty$ as $x \uparrow \infty$, $\int^\infty G(x)\,dx/x^2$ is convergent.*

Proof. In the corollary to Theorem 2, replace u_n by $a_n s_n$ and D_n by d_n/a_n, setting $g(x) = G(x)/x$.

Remarks. (1) The special choice of $g(x) \equiv G(x)/x$ in the corollary under Theorem 2, along with $d_n \equiv 1$, gives Rademacher's sequence of tests for convergence factors [**6**, Satz IV and addition thereto].

(2) The point of Theorem 5 (as compared with that of the first corollary to Theorem 3 in the case $a_n/d_n \downarrow 0$) consists in establishing a conclusion stronger than $\lim \sigma(s_n, a_n) = 0$ without assuming that $\lim \sigma(s_n, d_n) = 0$ but assuming instead a condition which in general implies less, namely that $\sigma(s_n, d_n)$ has 0 for one of its limit points. Suppose for instance the function $y = G(x)$ has a positive continuous derivative. Its inverse $x = G^{-1}(y)$ is then an increasing continuous function of y such that

$$\int^\infty dy/G^{-1}(y) = \int^\infty (G'(x)/x)\,dx = \{G(x)/x\}^\infty + \int^\infty (G(x)/x^2)\,dx$$

is convergent and therefore also $\sum d_n/G^{-1}(D_n)$ is convergent [**8**, p. 398, ex. 3]. In these circumstances, the condition $|\sigma(s_n, d_n)| \leq CG(d_n/a_n)/D_n$ assumed in Theorem 5 implies a limit point at 0 for $\sigma(s_n, d_n)$, since otherwise we must have, for all sufficiently large n,

$$CG(d_n/a_n)/D_n > \delta > 0$$

i. e.,

$$d_n/a_n > G^{-1}(D_n\delta/C) \text{ or } a_n < d_n/G^{-1}(D_n\delta/C)$$

which gives a contradiction (in consequence of what we have proved about G^{-1}) when, as in Theorems 3, 4, we postulate the divergence of $\sum a_n$.

4. I conclude with some examples which show how certain types of behaviour of $\sum \pm a_n$ affect or are affected by $\sum \pm d_n$, with the same distribution of signs, when we assume that a_n/d_n is monotonic. Of these

examples, I and II which were given by Cesàro in the case $d_n \equiv 1$ [3, §§ 1, 2] are obtained by taking $s_n = \pm 1$ in Theorems 3 (i) and 4.

I. *If a_n/d_n is decreasing, a necessary condition for* $\sum_1^n \epsilon_\nu a_\nu (\epsilon_\nu \equiv \pm 1) = o(A_n)$ *is* $\underline{\lim}\,(D^+_n/D^-_n) \leq 1 \leq \overline{\lim}\,(D^+_n/D^-_n)$ *where D^+_n and D^-_n denote respectively the sums of the d_ν's in* $\sum_1^n \epsilon_\nu d_\nu$, *with positive and negative coefficients, so that* $D^+_n + D^-_n = D_n$.

II. *If $a_n/d_n \downarrow 0$ and* $\underline{\lim}\,(a_n D_n/d_n) > 0$, *a necessary condition for the convergence of* $\sum \epsilon_\nu a_\nu$ *is* $\lim\,(D^+_n/D^-_n) = 1$.

III. *Let a_n be defined by*

$$a_n = d_n \frac{\prod_{\nu=1}^{n} (\alpha d_\nu + D_{\nu-1})(\beta d_\nu + D_{\nu-1})}{\prod_{\nu=1}^{n} (\gamma d_\nu + D_{\nu-1})(\delta d_\nu + D_{\nu-1})}$$

where $\{D_n\}$ is an increasing divergent sequence such that $d_n = D_n - D_{n-1} = O(1)$ *and $\alpha, \beta, \gamma, \delta$ are real numbers of which the two latter are different from* $-D_{\nu-1}/d_\nu$ $(\nu = 1, 2, 3, \cdots)$. *Then, for* $\mu = \gamma + \delta - \alpha - \beta > 1$, $\sum \epsilon_n a_n$ *is absolutely convergent; while for* $\mu \leq 1$, $\lim\,(\sum_1^n \epsilon_\nu a_\nu)/A_n = (l-1)/(l-1)$ *provided* $l = \lim\,(D^+_n/D^-_n)$ *exists.*

The proof starts with the fact

$$\frac{a_n/d_n}{a_{n+1}/d_{n+1}} = 1 + \mu(d_{n+1}/D_n) + O(d_{n+1}/D^2_n)$$

which, as shown elsewhere [7], is equivalent to $a_n \sim A d_n/D_n^{\mu}$, $A \neq 0$, and employs Theorem 3 (i) or 3 (ii) according as $0 < \mu \leq 1$ or $\mu < 0$; the proof of the case $\mu = 0$ being direct.

An interesting special case is that which $d_n \equiv \delta = 1$ when the series $\sum a_n$ reduces to the hypergeometric type.

IV. *Let a_n be defined in terms of d_n as in* III. *Then, for* $0 < \mu \leq 1$, $\sum \epsilon_n a_n$ *is convergent provided* $D^+_n - D^-_n = O\{G(D_n^{\mu})\}$ *where $G(x)$ is either* $x^{1-\rho}$, $0 < \rho < 1$, *or* $x/L_k(x)\,(\log_k x)^{\rho}$, $\rho > 0$.

This is a deduction from Theorem 5. As an example of it we notice that if $d_n = e^{1/n^2}$, $\sum e_n a_n$ is convergent provided $p_n - q_n = O\{G(n^{\mu})\}$, $0 < \mu \leq 1$, where p_n and q_n are the numbers of positive and negative terms

respectively in the n-th partial sum of $\sum \epsilon_n a_n$. It may be remarked that in this case the familiar Abel-Dirichlet test tells us nothing more than that $p_n - q_n = O(1)$ secures the convergence of $\sum \epsilon_n a_n$.

MADRAS CHRISTIAN COLLEGE,
TAMBARAM, S. INDIA.

REFERENCES

1. W. H. J. Fuchs, "A note on convergence factors," *Proceedings of the Edinburgh Mathematical Society*, (2) vol. 7 (1942), pp. 27-30.

2. G. H. Hardy, "On certain oscillating series," *Quarterly Journal of Mathematics*, vol. 38 (1907), pp. 269-288.

3. G. H. Hardy, "Note on a theorem of Cesàro," *Messenger of Mathematics*, vol. 41 (1912), pp. 17-22.

4. J. Karamata, "Über einen Konvergenzsatz des Herrn Knopp," *Mathematische Zeitschrift*, vol. 40 (1936), pp. 421-425.

5. J. Karamata, "Über einige reihentheoretische Sätze," *Mathematische Zeitschrift*, vol. 41 (1936), pp. 67-74.

6. H. Rademacher, "Über die asymptotische Verteilung gewisser Konvergenzerzeugender Factoren," *Mathematische Zeitschrift*, vol. 11 (1921), pp. 276-288.

7. C. T. Rajagopal, "Remarks on some generalizations of Cauchy's condensation and integral tests," *American Mathematical Monthly*, vol. 48 (1941), pp. 180-185.

8. C. de la Vallée-Poussin, *Cours d'Analyse*, 1 (Louvain-Paris, 1926).

ON MONOTONE SERIES.*

By Aryeh Dvoretzky.

1. Introduction. Throughout the paper c_n $(n=1,2,\cdots)$ denotes the n-th term of a monotone convergent series of positive terms and d_n $(n=1,2,\cdots)$ the n-th term of a monotone divergent series of positive terms. A. Pringsheim ([4],[1] explicitly [5], p. 615; see also e.g. [1], p. 47, [3], p. 303) pointed out that the inequality

$$c_n \geq d_n \tag{1}$$

may be realized for infinitely many n.

Let $N(m)$ denote the number of n among $1,2,\cdots,m$ for which (1) holds and put $A(m)=N(m)/m$. Recently, R. W. Hamming [2] showed the rarity, in a certain sense, of the occurrence of (1) by proving

$$\liminf A(n) = 0.^2 \tag{2}$$

On the other hand he constructed an example for which

$$\limsup A(n) = 1. \tag{3}$$

The relation (2) is an immediate corollary of the following result of Hamming:

For every $\epsilon > 0$ and $k > 1$ there exist infinitely many r_i, $r_i \to \infty$, such that for all n with $r_i \leq n \leq kr_i$, we have $A(n) < \epsilon$.

Using a very simple argument we establish (in **2**) the following

Theorem 1. *For every $K > 1$ there exist infinitely many R_i, $R_i \to \infty$, such that for all n with $R_i \leq n \leq KR_i$, we have $c_n < d_n$.*

This theorem clearly implies Hamming's result, since for $r_i = R_iK^{\frac{1}{2}} \leq n \leq KR_i = kr_i$ we have $A(n) < R_i/(R_i + R_iK^{\frac{1}{2}} - 1) < K^{-\frac{1}{2}}$.

Let $P(m)$, $0 \leq P(m) \leq \infty$, be the number of consecutive indices n beginning with $n=m$ for which $c_n < d_n$. Theorem 1 is evidently equivalent to

$$\limsup P(n)/n = \infty. \tag{4}$$

* Received March 15, 1947.

[1] Numbers in square brackets refer to the bibliography at the end of the paper.

[2] All limits are taken for $n \to \infty$. Similarly, summations are (unless otherwise stated) upon all positive integers.

Similarly, Hamming's result is equivalent to

$$\limsup Q(n;\epsilon)/n = \infty \text{ for every } \epsilon > 0, \tag{5}$$

where $Q(m;\epsilon)$, $0 \leq Q(m;\epsilon) \leq \infty$, denotes the number of consecutive indices n beginning with $n = m$ for which $A(n) < \epsilon$.

After some preparations we show (in **5**) by means of suitable examples that (2) as well as Hamming's result and Theorem 1 are the best possible general results of their kind. More precisely, we prove

THEOREM 2. *Let g_n $(n = 1, 2, \cdots)$ be any sequence of positive numbers tending (however slowly) to infinity. Then there exist series Σc_n and Σd_n for which*

$$\lim g_n A(n) = \infty, \tag{6}$$

$$\lim P(n)/(ng_n) = 0, \tag{7}$$

$$\lim Q(n; 1-\epsilon)/(ng_n) = 0, \qquad \textit{for every } \epsilon > 0. \tag{8}$$

In the opposite direction we improve on (3) by establishing (in **4**)

THEOREM 3. *Let G_n $(n = 1, 2, \cdots)$ be any sequence of positive numbers satisfying*

$$\liminf G_n/n = 0. \tag{9}$$

Then there exist series Σc_n and Σd_n for which

$$\liminf G_n(1 - A(n)) = 0. \tag{10}$$

This result is also the best of its kind. In fact the requirement (9) cannot be replaced by a less stringent one, since $n(1 - A(n)) = n - N(n)$ is the number of terms up to the n-th for which $c_\nu < d_\nu$, and thus increases to infinity with n.

2. Proof of Theorem 1. The theorem is obvious when the d_n are non-decreasing, hence we may assume them non-increasing. Also no loss of generality is involved in supposing K an integer ≥ 2.

Put

$$C_n = \sum_{K^{n-1}}^{K^n-1} c_\nu, \qquad D_n = \sum_{K^{n-1}}^{K^n-1} d_\nu \qquad (n = 1, 2, \cdots).$$

If $c_\nu \geq d_\nu$ for some ν with $K^{n-1} \leq \nu \leq K^n$, $(n > 1)$ then, since both series are

monotone, it follows that every term of C_{n-1} is not smaller than every term of D_{n+1}, whence

$$C_{n-1}/(K^{n-1}-K^{n-2}) \geq D_{n+1}/(K^{n+1}-K^{n}),$$

or

$$C_{n-1} \geq D_{n+1}/K^{2}.$$

But ΣC_n is convergent and ΣD_n is divergent, hence there must be infinitely many n for which the last inequality does not hold. Thus there are infinitely many n for which $c_\nu < d_\nu$ throughout the range $K^{n-1} \leq \nu \leq K^n$. This is precisely the contention of Theorem 1.

3. A Lemma. We need the following lemma which is of some interest by itself.

LEMMA. *Let f_n $(n = 1, 2, \cdots)$ be a monotone sequence of positive numbers tending (however slowly) to zero. Then there exist series Σc_n and Σd_n with $c_n = f_n d_n$.*

If we do not require the monotone character of Σc_n and Σd_n, then the lemma reduces to a classical remark of Stieltjes (see e. g. [1], p. 47 or [3], p. 302).[3]

Proof. Let $n_1 \geq 1$ be such that $f_{n_1} < 2^{-1}$. Let $n_2 > 2n_1$ be such that $f_{n_2} < 2^{-2}$, and generally let (for $k = 2, 3, \cdots$) $n_k > 2n_{k-1}$ be such that $f_{n_k} < 2^{-k}$. Putting $n_0 = 0$ we define $d_\nu = 1/(n_k - n_{k-1})$ for $n_{k-1} < \nu \leq n_k$, $(k = 1, 2, \cdots)$. Clearly $d_\nu \geq d_{\nu+1}$ and $\Sigma d_\nu = \infty$. Evidently $c_\nu = f_\nu d_\nu$ is also monotone and, since $c_{n_k+1} + \cdots + c_{n_{k+1}} < 2^{-k}$, the series Σc_ν is convergent, and the lemma proved.

We remark that the demonstration yields (also for Σd_n) monotone non-increasing series.

4. A construction. Let t_n $(n = 1, 2, \cdots)$ be any monotone increasing sequence of positive integers satisfying

$$\lim t_n/t_{n+1} = 0. \tag{11}$$

[3] We are unacquainted with an explicit formulation of the lemma and the nearest published results seems to be a statement of Pringsheim [5, pp. 609-10] where our $c_n = f_n d_n$ is replaced by $c_n \leq f_n d_n$. For proof Pringsheim refers to [4], but his statement does not appear to be a corollary of any of the theorems of [4], though it can of course be derived by arguments similar to those used in that paper. In fact, we deduce the lemma by means of a construction similar to some frequently employed by Pringsheim and others in related investigations.

The lemma assures the possibility of constructing a sequence of positive numbers a_n $(n=1,2,\cdots)$ such that both series $\Sigma a_n t_{2n-1}$ and $\Sigma a_n t_{2n}$ are monotone non-increasing, the first being convergent the other divergent. Since $a_n t_{2n-1} \geq a_{n+1} t_{2n+1}$ we have also $a_n > a_{n+1}$, i. e. the a_n decrease steadily to zero.

Putting $T_{-1}=T_0=0$ and $T_n = t_1 + t_2 + \cdots + t_n$ for $n=1,2,\cdots$, we define

$$(12)\qquad \begin{aligned} c_\nu &= a_n \quad \text{for} \quad T_{2n-3} < \nu \leq T_{2n-1} \qquad (n=1,2,\cdots), \\ d_\nu &= a_n \quad \text{for} \quad T_{2n-2} < \nu \leq T_{2n} \qquad (n=1,2,\cdots). \end{aligned}$$

Clearly, (c_ν) and (d_ν) are monotone sequences. Furthermore, since $T_{2n-1} - T_{2n-3} < 2t_{n-1}$ and $T_{2n} - T_{2n-2} > t_{2n}$ we have $\Sigma c_\nu < \infty$ and $\Sigma d_\nu = \infty$ as required.

We note that

$$c_\nu = d_\nu \quad \text{for} \quad T_{2n-2} < \nu \leq T_{2n-1} \qquad (n=1,2,\cdots)$$

while for other values of ν we have $c_\nu < d_\nu$.

Thus $N(m)$ and $A(m)$ (see **1**) increase as m increases from T_{2n-2} to T_{2n-1} and decrease as m increases from T_{2n-1} to T_{2n}.

In particular, we have

$$(13)\qquad \begin{aligned} A(T_{2n-1}) &= (t_1 + t_3 + \cdots + t_{2n-1})/T_{2n-1} \geq t_{2n-1}/T_{2n-1} \\ A(T_{2n}) &= (t_1 + t_3 + \cdots + t_{2n-1})/T_{2n} \geq t_{2n-1}/T_{2n}. \end{aligned}$$

Also

$$(14)\qquad P(m) \leq T_{2n} - T_{2n-1} = t_{2n} \quad \text{for} \quad T_{2n-1} < m \leq T_{2n+1}.$$

Furthermore, for given $\epsilon > 0$ we have

$$(15)\qquad Q(m; 1-\epsilon) \leq T_{2n-1}/\epsilon \quad \text{for} \quad T_{2n-2} < m \leq T_{2n},$$

provided $T_{2n-1}/\epsilon < t_{2n}$; which, in view of (11), holds for n sufficiently large since (11) entails

$$(16)\qquad \lim t_n/T_n = 1.$$

5. Proof of Theorem 2. We may, without loss of generality, assume the g_n to be non-decreasing and, furthermore, each g_n to be the square of a positive integer greater than unity.

The sequence

$$(17)\qquad t_1 = 1, \qquad t_{n+1} = t_n g_n^{\frac{1}{2}} \qquad (n=1,2,3,\cdots)$$

is monotone increasing and satisfies (11). Hence we may construct series Σc_ν and Σd_ν as in the last section.

By (13), the remark preceding it and (17), we have for $T_{2n-1} < m \leq T_{2n+1}$

$$g_m A(m) \geq g_m A(T_{2n}) \geq g_m t_{2n-1}/t_{2n} = g_m/g_{2n-1}{}^{\frac{1}{2}}.$$

Since $m \geq T_{2n-1} > 2n-1$ for $n > 1$, the last member of the above inequality is greater than $g_{2n-1}{}^{\frac{1}{2}}$ and thus (6) holds.

Again, taking account of (14), we have for $T_{2n-1} < m \leq T_{2n+1}$

$$P(m)/(mg_m) < t_{2n}/(T_{2n-1}g_m) \leq g_{2n-1}{}^{\frac{1}{2}}t_{2n-1}/(t_{2n-1}g_m) < g_{2n-1}{}^{-\frac{1}{2}}$$

and thus (7) is satisfied.

Finally, by (15) and the remark following it, we have for $T_{2n-2} < m \leq T_{2n}$ and n sufficiently large

$$Q(m; 1-\epsilon)/(mg_m) < T_{2n-1}/(\epsilon g_m T_{2n-2});$$

which, by (17), (16) and $m > 2n-2$, yields (8), thus completing the proof of Theorem 2.

6. Proof of Theorem 3. We construct a sequence T_n of increasing integers as follows: we take $T_1 = 1$ and, having defined T_ν for $\nu = 1, 2, \cdots, n-1$, we choose as T_n any integer satisfying the conditions

$$T_n > nT_{n-1}, \qquad nT_{n-1}G_{T_n} < T_n; \tag{18}$$

the existence of an integer satisfying these conditions is assured by (9).

Clearly, the sequence

$$t_1 = T_1, \qquad t_n = T_n - T_{n-1} \qquad (n = 2, 3, \cdots)$$

satisfies (11) and we proceed with the construction of Section **4.**

Then, by (13) and (18), we have for $m = T_{2n-1}$,

$$G_m(1 - A(m)) \leq G_m T_{2n-2}/T_{2n-1} < 1/(2n-1),$$

and thus (10) holds.

7. Remarks. 1. The series Σc_n and Σd_n constructed in the paper are monotone non-increasing; however an easy modification renders them strictly decreasing. In fact, let $\Sigma \gamma_n$ be a convergent strictly monotone series of positive terms and replace the c_n and d_n of Theorems 2 and 3 by $c_n + \gamma_n$ and $d_n + \gamma_n$ respectively (and the d_n of the lemma by $d_n + \gamma_n$).

2. Hamming's result (5) is only apparently stronger than (2). In fact, $A(n) < \delta$ implies $A(\nu) < k\delta$ for $n \leq \nu \leq kn$. Since for given k we can, granted (2), take $\delta < \epsilon/k$, and n arbitrarily large, (5) follows.

3. (7) can be deduced directly from (6). In fact as Theorem 2 gains in strength when the growth of g_n becomes slower, we may assume $g_m/g_n \to 1$ provided $m, n \to \infty$ so that $m/(ng_n) \to 1$. If $P(n) > \delta n g_n$ then $A(n + \delta n g_n) < n/(n + \delta n g_n) < 1/(\delta g_n)$. Thus, in view of the assumption on the growth of g_n, the failure of (7) implies that of (6).

4. Pringsheim[4] showed not only that (1) may be realized an infinity of times but that moreover we may have $\liminf d_n/c_n = 0$. If we define $N_H(m)$ as the number of n among $1, 2, \cdots, m$ for which $c_n \geq H d_n$ and similarly for A_H, P_H and Q_H it can be shown that the relations (6), (7), (8) of Theorem 2 and (10) of Theorem 3 hold *for every* $H > 0$ *with the same series* Σc_n *and* Σd_n when A is replaced by A_H etc. In fact all that is needed is to replace the definition of d_ν in (12) by $d_\nu = a_n/s_n$ where s_n increases to infinity and so slowly that $\Sigma a_n t_{2n}/s_n$ is still divergent. One may take e. g.

$$s_n = a_n t_{2n} / \sum_1^n a_\nu t_{2\nu}.$$

5. Pringsheim[4] actually showed more. He proved that given a series Σc_n there exists a series Σd_n with $\liminf d_n/c_n = 0$ and vice versa. In Theorems 2 and 3 we made essential use of the possibility of choosing *both* series. If one of them is given the theorems must be replaced by weaker ones. Thus, if the divergent series is $\Sigma 1/n$ we have $N(m)/m \to 0$ and even (3) fails.

6. Hamming's result on summability [2], p. 136 is also an immediate corollary of Theorem 1.

7. All the results of the paper carry over, *mutatis mutandis,* to infinite integrals of positive monotone functions.

8. Extensions to double series etc., though necessitating much preliminary work, also do not present new difficulties. Various extensions are possible corresponding to the various manners of defining monotone series and convergence.

EINSTEIN INSTITUTE OF MATHEMATICS,
HEBREW UNIVERSITY, JERUSALEM.

[4] Same references as in the beginning of 1.

BIBLIOGRAPHY

[1]. T. J. I'A. Bromwich, *An Introduction to the Theory of Infinite Series*, second ed., London, 1942.

[2]. R. W. Hamming, "Monotone Series," *American Journal of Mathematics*, vol. 68 (1946), pp. 133-136.

[3]. K. Knopp, *Theory and Application of Infinite Series*, London and Glasgow, 1928.

[4]. A. Pringsheim, "Allgemeine Theorie der Divergenz und Konvergenz von Reihen mit positiven Gliedern," *Mathematische Annalen*, vol. 35 (1890), pp. 296-394.

[5]. ———, "Ueber die sogenannte Grenze und die Grenzgebiete zwischen Convergenz und Divergenz," *Sitzungsberichte der mathematisch-physikalischen Classe der k. b. Akademie der Wissenschaften zu München*, vol. 36 (1896), pp. 605-624.

THE ISEPIPHAN PROBLEM FOR n-HEDRA.*

By LÁSZLÓ FEJES TÓTH.

1. Following T. Bonnesen[1] we call the analogon in space of the isoperimetric problem the isepiphan problem. The problem is to pick out from a class of isepiphan polyhedra, i. e., polyhedra with equal surface areas F, that one with greatest volume V.

We obtain different problems by restricting the class of polyhedra from which we want to pick out the best one, i. e. the one with the smallest value of F^3/V^2, to polyhedra

a) with a given number of faces,

b) with a given number of vertices,[2]

c) of a given type (i. e. to topologically isomorphic polyhedra), or to polyhedra upon which any other condition is imposed.

E. Steinitz[3] gives a detailed account of the history of this range of problems. Let us mention only the conjecture of J. Steiner according to which the regular polyhedra are the best in the class of their own type. The difficulties involved in these problems are illustrated by the fact that Steinitz was not able to decide the correctness of this conjecture even for the case of the cube, and the conjecture concerning the icosahedron and the dodekahedron is considered by him as entirely unfounded.[4]

In this paper we prove the following

* Received September 19, 1947.

[1] T. Bonnesen, *Sur les problèmes des isoperimètres et des isepiphanes*, Paris, 1929, Bonnesen-Fenchel, *Theorie der konvexen Körper*, Berlin, 1934.

[2] Concerning the case b) we mention only that the best polyhedron is for any number of vertices bounded only by triangles (see the second paper in [5]). This fact is the dual counterpart of Goldberg's conjecture (see note [5]) according to which the best polyhedron among the polyhedra with a given number of faces, has only trihedral vertices.

[3] E. Steinitz, "Polyeder und Raumeinteilungen," *Encyklopädie der Mathematischen Wissenschaften*, III-12, 9 (1922), pp. 38-43, 94-101; "Über isoperimetrische Problem bei konvexen Polyedern," *Journal für die reine und angewandte Mathematik*, vol. 158 (1927), pp. 129-153, 159 (1928), pp. 133-143.

[4] See the second paper cited in [3], beginning of § 4. resp. end of § 3. on p. 134. In contradiction to the above remark Goldberg mentions the cube among the polyhedra which are proved to be the best in the case a).

THEOREM 1. *If F_n denotes the area and V_n the volume of any convex n-hedron, we have*

$$(1) \qquad F_n^3/V_n^2 \geqq 54(n-2)\tan\omega_n\,(4\sin^2\omega_n - 1), \qquad \omega_n = \frac{n}{n-2}\,\frac{\pi}{6}.$$

Equality holds only for the regular tetrahedron, hexahedron and dodecahedron.[5]

According to this, the regular hexahedron and dodecahedron are proved to be the best not only among those of their type but also—far beyond Steiner's conjecture—among all convex polyhedra with 6 resp. 12 faces.

The above inequality gives furthermore an *exact asymptotical estimation for large values of n.* The only lower bound of F_n^3/V_n^2 for large n previously known was 36π, the value of F^2/V^2 for the sphere. The above estimation gives besides that the exact value of the constant $\liminf n(F_n^3/V_n^2 - 36\pi)$.

By the theorem indicated above the isepiphan problem for n-hedra (case a)) can in a certain sense be considered as closed since the determination of the exact lower bound for all values of n seems unattainable or must be considered at least as a farther aim.

2. According to a remarkable result of L. Lindelöf and Minkowski[6] among all convex n-hedra with given directions of the exterior normals of the faces, the n-hedron circumscribed about a sphere is the best. Therefore we can restrict ourselves to n-hedra circumscribed about the sphere S of unit radius. We have then $V_n = \frac{1}{3}F_n$, i. e. $F_n^3/V_n^2 = 9F_n = 27V_n$, and thus the isepiphan problem for n-hedra is equivalent to the problem of determining the n-hedron with minimal area (or volume) among all n-hedra circumscribed about the sphere of unit radius.

---

[5] This result was previously obtained by M. Goldberg [" The isoperimetric problem for polyhedra," *Tôhoku Mathematical Journal*, vol. 40 (1935), pp. 226-236]. Goldberg's proof is based upon the convexity of a complicated function of two variables which fact he states without proof. Taking into account the importance of this result it seems worth while to give a complete proof.—In a paper [" Über einige Extremaleigenschaften der regulären Polyeder und des gleichseitigen Dreiecksgitters," *Annali della R. Scuola Normale Superiore di Pisa* 12 (1943) . . .] I have proved independently from Goldberg but in a similar way—together with a number of analogous inequalities—the inequality (1). In this paper there can be found also a nomogram of the function in question from which the convexity can be seen empirically.

[6] L. Lindelöf, " Propriétés générales des polyedres etc.," *St. Petersburg Bull. Ac. Sc.*, vol. 14 (1869), pp. 258-269. H. Minkowski, " Allgemeine Lehrsätze über die konvexen Polyeder," *Gesammelte Abhandlungen*, vol. 2, pp. 103-121.

All that we have to prove is, therefore, the inequality

$$F_n \geq 6(n-2) \tan \omega_n (4 \sin^2 \omega_n - 1)$$

concerning an n-hedron circumscribed to the unit sphere S.

We prove this inequality as a corollary of the following

THEOREM 2. *Let $\phi(\rho)$ be a strictly increasing function for $0 \leq \rho < \pi/2$. Let further $P_1, P_2, \cdots, P_n$ be n points of the surface S of the sphere with unit radius not all of them lying on a hemisphere, and let $\rho_P = \min(PP_1, PP_2, \cdots, PP_n)$ denote the spherical distance of a variable point P of S from the point of the system $\{P_i\}$ nearest to it.*

With these conditions we have for the surface-integral of $\phi(\rho_P)$ extended over S

$$\int_S \phi(\rho_P)\, d\omega \geq (2n-4) \int_{\bar{\Delta}} \phi(\bar{\rho}_P)\, d\omega, \tag{2}$$

where we denote by $\bar{\Delta}$ an equilateral spherical triangle with the vertices $\bar{P}_1, \bar{P}_2, \bar{P}_3$ having the area[7] *$2\pi/(n-2)$ and by $\bar{\rho}_P = \min(P\bar{P}_1, P\bar{P}_2, P\bar{P}_3)$ the spherical distance of P from the vertex nearest to it.*

Equality holds only if the system $\{P_i\}$ coincides with the system of vertices of a regular tetrahedron, octahedron or icosahedron.

First we make two preliminary remarks, denoting in what follows a domain and its area by the same symbol.

Remark 1. Let s be a segment of a spherical cap c with the top point T. We assert that the function

$$\psi(s) = \int_s \phi(PT)\, d\omega$$

is convex from above for $0 \leq s \leq c/2$.

Proof. Let s_1, s^*_1, s_2, s^*_2 be four segments of c cut off by great circles passing through the same diametrically opposite points, for which

---

[7] We can suppose that the least convex envelope V_n of $P_1, P_2, \cdots, P_n$ is a polyhedron bounded only by triangles. The number of these triangles is then $2n-4$. Hence $4\pi/(2n-4)$ is the average area of the spherical triangles determined by the spherical net of V_n.

$0 < s_1 < s^*_1 < s_2 < s^*_2 < c/2$ and $\Delta s_1 = s^*_1 - s_1 = \Delta s_2 = s^*_2 - s_2$. It is easy to give an area-preserving representation of Δs_1 upon Δs_2 so that for any point P_2 of Δs_2 corresponding to an inner point P_1 of Δs_1 we have $TP_1 > TP_2$. Hence by the monotoneity of $\phi(\rho)$

$$\int_{\Delta s_1} \phi(PT)\, d\omega > \int_{\Delta s_2} \phi(PT)\, d\omega$$

and thus $\psi'(s_1) > \psi'(s_2)$.

Remark 2. For all convex domains d lying in a hemicap of c

$$\int_d \phi(PT)\, d\omega \leq \psi(d).$$

Proof. Let d be incongruent to s. Denote the point d nearest to T by D, the surface of d resp. s lying in the concentric cap with radius ρ by $d(\rho)$ resp. $s(\rho)$. It is shown easily that for $TD < \rho < \bar{\rho}$, $s(\rho) > d(\rho)$, where ρ denotes the radius of c. This involves an area-preserving representation of s upon d which carries all inner points of s into points of d nearer to T.

Let us now turn to the

Proof of Theorem 2. Consider the n-hedron U_n the faces of which touch S at $P_1, P_2, \cdots, P_n$. Obviously we can suppose that U_n has only trihedral vertices, since a ν-hedral vertex can be considered as the limit-position of $\nu - 2$ trihedral vertices. The number of the edges of U_n (some of which can shrink to a point) is then exactly $3n - 6$.

Let f_i be the face with the tangent point P_i, ϕ_i the projection of f_i from the centre O of S upon S. Obviously we have

$$\int_S \phi(\rho_P)\, d\omega = \sum_{r=1}^{n} \int_{\phi_i} \phi(PP_i)\, d\omega.$$

Let c_i be the spherical cap with the top point P_i and the spherical radius $\bar{\rho}$ congruent with the least cap c with the top point T containing $\bar{\Delta}$. Let, further, $s_1, s_2, \cdots, s_\nu$ be the convex partial-domains of c_i the first bordered by the great-circles P_iQ_1, Q_1Q_2, P_iQ_2, the second by P_iQ_2, Q_2Q_3, P_iQ_3, etc., $Q_1, Q_2, \cdots, Q_\nu$ being the vertices of the spherical polygon ϕ_i. We have

$$\int_{\phi_i} \phi(PP_i)\, d\omega = \int_{c_i} \phi(PP_i)\, d\omega - \sum_{r=1}^{\nu} \int_{s_r} \phi(PP_i)\, d\omega + \int_{\phi_i} \phi(PP_i)\, d\omega,$$

where ϕ_i denotes the part of ϕ_i not covered by c_i. The total number of the

domains s_r being $6n-12$—i. e. twice the number of the edges of U_n—we get by summation

$$\int_S \phi(\rho_P)\,d\omega = n \int_c \phi(PT)\,d\omega - \sum_{r=1}^{6n-12} \int_{s_r} \phi(PT)\,d\omega + \int_{S'} \phi(\rho_P)\,d\omega,$$

where S' denotes the part of S not covered by any cap c_i and the $s_\mu - s$ are parts of c congruent with the original domains characterized above.

Taking now into account that

$$S = nc - \sum_{r=1}^{6n-12} s_r + S'$$

we have by Remark 2. resp. 1 and by Jensen's inequality,[8]

$$\sum_{r=1}^{6n-12} \int_{s_r} \phi(PT)\,d\omega \leq \sum_{r=1}^{6n-12} \psi(s_r) \leq (6n-12)\psi\left(\frac{\sum_{r=1}^{6n-12} s_r}{6n-12}\right)$$
$$= (6n-12)\psi\left(\frac{nc-S+S'}{6n-12}\right)$$
$$= (6n-12)\psi\left(\frac{nc-S}{6n-12}\right) + (6n-12)\int_d \phi(PT)\,d\omega.$$

Here we have denoted by d the domain with the area $d = S'/(6n-12)$ which completes the segment of the cap c of area $(nc-S)/(6n-12)$ to the segment of area $(nc-S+S')/(6n-12)$.

For the inner points P of d resp. of S'—provided that they are not empty—$PT < \bar{\rho}$ resp. $\rho_P > \bar{\rho}$ and thus by the monotoneity of $\phi(\rho)$,

$$(6n-12)\int_d \phi(PT)\,d\omega \leq \int_{S'} \phi(\rho_P)\,d\omega.$$

Taking all these facts into account, we have

$$\int_S \phi(\rho_P)\,d\omega \geq n \int_c \phi(PT)\,d\omega - (6n-12)\psi\left(\frac{nc-S}{6n-12}\right).$$

Consider now the integral $\int_{\bar{\Delta}} \phi(\bar{\rho}_P)\,d\omega$. Let us denote the angles of $\bar{\Delta}$ by α. We denote by $3s$ the part of $\bar{\Delta}$ which is covered twice by the caps having the top points $\bar{P}_1$, $\bar{P}_2$, $\bar{P}_3$, and congruent with c. We have

$$\Delta = 3(\alpha/2\pi)c - 3s,$$

[8] J. L. W. V. Jensen, "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta Mathematica*, vol. 30 (1906), pp. 175-193.

and analogously

$$\int_{\bar{\Delta}} \phi(\bar{\rho}_P)\,d\omega = 3(\alpha/2\pi)\int_c \phi(PT)\,d\omega - 3\psi(s).$$

Thus since $3\alpha - \pi = \bar{\Delta}$,

$$\int_{\Delta} \phi(\bar{\rho}_P)\,d\omega = \left\{ (\bar{\Delta}+\pi)/2\pi \right\} \int_c \phi(PT)\,d\omega - 3\psi\left(\frac{\bar{\Delta}c + \pi c - 2\pi\bar{\Delta}}{6\pi}\right).$$

Now we make use of the supposition $\bar{\Delta} = 2\pi/(n-2)$ and obtain

$$(2n-4)\int_{\bar{\Delta}} \phi(\bar{\rho}_P)\,d\omega = n\int_c \phi(PT)\,d\omega - (6n-12)\psi\left(\frac{nc-4\pi}{6n-12}\right),$$

which equals the right side of the last inequality.

Equality holds in (2) only if S is entirely covered by the caps c_i without any part being covered three times, the $3n-6$ parts covered twice being all congruent. In this case U_n is a regular polyhedron with trihedral vertices; this completes the proof.[9]

3. Let ϕ be a spherical domain lying on a hemisphere with the top point T, f the central-projection of ϕ on the plane touching the sphere S at T. It is easy to see that f is given by an integral

$$f = \int_{\phi} \phi(PT)\,d\omega,$$

$\phi(\rho)$ being strictly increasing[10] for $0 \leq \rho < \pi/2$.

For this function the left side of (2) gives the area of U_n, the right side the area of the projection of $\bar{\Delta}$ upon the trihedron, the faces of which touch S at the vertices of $\bar{\Delta}$. Hence by some elementary computations we obtain the desired estimation for F_n.

Similarly we get for instance from Theorem 2 the following

COROLLARY. *Let $S_1, S_2, \cdots, S_{12}$ be 12 congruent spheres all of them containing the fixed sphere S. The convex body B consisting of the common*

[9] Let us mention here the following interesting fact: The surface-integral

$$\int_{\Delta} \phi(\rho_p)\,d\omega; \qquad \rho_p = \min(PP_1, PP_2, PP_3)$$

extended over the spherical triangle $\Delta \equiv P_1P_2P_3$ of given area takes its minimum 1 for an equilateral triangle $\bar{P}_1\bar{P}_2\bar{P}_3$, 2 for a triangle $\bar{P}_1\bar{P}_3\bar{P}^*_1$ arising from $\bar{P}_1\bar{P}_2\bar{P}_3$ by reflecting $\bar{P}_1$ with respect to $\bar{P}_2\bar{P}_3$ in $\bar{P}^*_1$, provided that the height of $\bar{P}_1\bar{P}_2\bar{P}_3$ is $< \pi/2$.

[10] We have $\phi(\rho) = 1/\cos^3\rho$.

points of the spheres has minimal area and volume if the spheres touch S at the vertices of a regular icosahedron.

Finally let us mention that in (2) the case of equality remains unaltered for the function

$$\phi(\rho) = \begin{cases} 0 \text{ for } 0 \leq \rho \leq \bar{\rho} \\ 1 \text{ for } \bar{\rho} < \rho < \pi/2. \end{cases}$$

This shows that S cannot be covered by n congruent caps of a spherical radius $< \bar{\rho}$.

The result obtained can be formulated as follows: If d denotes the density of any system of $n > 2$ congruent spherical caps of areas c, covering the unit sphere, then

$$d = nc/4\pi \geq n/2 \left(1 - (\sqrt{3}/3) \cot \frac{n}{n-2} \frac{\pi}{6}\right) > 2\sqrt{3}\pi/9 = 1 \cdot 209 \cdots$$

holds. This proposition found some time ago [11] is an extension of a result due to R. Kershner.[12]

BUDAPEST.

[11] L. Fejes, "Egy gömbfelület befedése egybevágó gömbsüvegekkel," *Mathematikai és phisikai lapok* 50 (1943), pp. 40-46.

[12] R. Kershner, "The number of circles covering a set," *American Journal of Mathematics*, vol. 61 (1939), pp. 665-671.

THE MEAN OF A FUNCTION OF EXPONENTIAL TYPE.*

By A. R. Harvey.[1]

1. Introduction.

Definition. The function $f(z)$ is said to be of exponential type k if $f(z)$ is entire and

$$M(r) = O(e^{(k+\epsilon)r})$$

for every $\epsilon > 0$ and no $\epsilon < 0$, $M(r)$ being the maximum of $|f(z)|$ on the circle $|z| = r$.

It is easily seen that

$$k = \overline{\lim_{r\to\infty}}\, r^{-1} \log M(r),$$

and it can be shown [35; p. 41] that

$$k = \overline{\lim_{n\to\infty}}\, |f^{(n)}(z)|^{1/n}$$

for every z.

Definition. The p-th mean of $f(z)$ on a horizontal line, denoted by $M^p\{(x + iy)\}$, is defined:

$$M^p\{f(x+iy)\} = \overline{\lim_{T\to\infty}}\, (1/2T) \int_{-T}^{T} |f(x+iy)|^p\, dx.$$

The p-th mean of $f(z)$ at the integers, denoted by $M_i{}^p\{f(x)\}$, is defined:

$$M_i{}^p\{f(x)\} = \overline{\lim_{n\to\infty}}\, (2n+1)^{-1} \sum_{i=-n}^{n} |f(i)|^p.$$

It is known (cf. [13; p. 143-144], [2; p. 102], [27; vol. II, p. 35-36, nos. 201-202]) that if $f(z)$ is of exponential type k and

$$|f(x)| \leq B$$

for all x, then

$$|f(x+iy)| \leq Be^{k|y|},$$

and

$$|f'(x)| \leq kB.$$

* Received May 13, 1947.

[1] The author is indebted to Dr. R. P. Boas, Jr. for suggesting the subject of this paper, and to Dr. Boas and Professors L. Ahlfors and D. V. Widder for many helpful suggestions.

It has been proved also ([11], [13]) that if k is less than π, the boundedness of $f(z)$ at the integers implies the boundedness of $f(z)$ on the whole x-axis. Results analogous to the preceding ones have been obtained for functions of exponential type which are of class L^p on the x-axis ([24; p. 120, 128], [7; p. 276]). In this paper, the analogues for functions of exponential type with bounded means will be obtained.

The results to follow are not very interesting in the case where $f(x)$ is bounded. The class of functions having a bounded mean of the form $M^p\{f(x)\}$, however, includes the class of B^1 (Besicovitch) almost periodic functions ([4], [5], [6]). Boas [8, p. 13] has shown that there exists a function of exponential type which is B^1 almost periodic but unbounded on the x-axis; hence the class of functions under consideration is greater than the class of functions which are of exponential type and which are bounded on the x-axis.

2. The mean of $f(z)$ on a horizontal line. It has been pointed out that a function may be unbounded on the x-axis yet have a bounded mean. It is to be expected, however, that such a function cannot grow very rapidly. The following theorem establishes a limit on the rate of growth of a function of exponential type with bounded mean.

THEOREM 1. *If $f(z)$ is of exponential type k, and if*

$$M^p\{f(x)\} = A < \infty, \qquad (p > 0),$$

then

$$f(x) = O(|x|^{1/p}) \qquad as\ |x| \to \infty.$$

Proof. Assume that $k < \pi$, otherwise consider $f(z/k)$. By hypothesis, there exists a constant B such that

$$(1/2T) \int_{-T}^{T} |f(x)|^p \, dx < B$$

for all T. Then, for any positive integer n and any δ with $0 < \delta < \frac{1}{2}$,

$$(1/2n) \int_{n-\delta}^{n} |f(x)|^p \, dx \leq (1/2n) \int_{-n}^{n} |f(x)|^p \, dx < B,$$

so that

$$\int_{n-\delta}^{n} |f(x)|^p \, dx < 2Bn.$$

It follows that for some a_n, with $n - \delta \leq a_n \leq n$,

$$|f(a_n)|^p < 2Bn\delta^{-1} \leq 4Ba_n\delta^{-1}.$$

Consider any branch of the function,

$$g(z) = (z+1)^{-1/p} f(z+1).$$

It is regular and of type k in $R(z) \geqq 0$, and is bounded at the sequence $\{a_n - 1\}$. Hence, by a theorem of Duffin and Schaeffer [13; p. 142], $|g(x)|$ is bounded. This proves the theorem for the case where x approaches infinity through positive values.

The case where x approaches infinity through negative values is proved in a similar way.

THEOREM 2. *If $f(z)$ is of exponential type k, then*

$$M^p\{f(x+iy)\} \leqq e^{pk|y|} M^p\{f(x)\}, \qquad (p > 0).$$

Proof. Assume that $y \geqq 0$, the case, $y \leqq 0$, being handled in a similar way. $M^p\{f(x)\}$ is assumed to be finite, otherwise there is nothing to prove.

By Theorem 1,

$$\int_{-\infty}^{\infty} \frac{\log |f(t)|}{1+t^2}\, dt < \infty,$$

so that by a result due to Nevanlinna [22; p. 29-31],

$$\log |f(re^{i\theta})| \leqq (1/\pi) \int_{-\infty}^{\infty} \log |f(t)| \cdot \frac{r \cdot \sin\theta}{r^2 + t^2 - 2rt \cdot \cos\theta}\, dt + \frac{2cr \cdot \sin\theta}{\pi},$$

where

$$c = \lim_{r\to\infty} \int_0^{\pi} \log |f(re^{i\theta})| \sin\theta d\theta.$$

Using Theorem 1, it is easily seen that, under the hypothesis of this theorem, $c = \frac{1}{2}\pi$, hence,

$$\log |f(x+iy)|^p \leqq (1/\pi) \int_{-\infty}^{\infty} \frac{y}{(x-t)^2 + y^2} \log |f(t)|^p\, dt + pky$$

$$\leqq \log \{(1/\pi) \int_{-\infty}^{\infty} \frac{y}{(x-t)^2 + y^2} |f(t)|^p\, dt\} + pky.$$

The last inequality is a consequence of the fact that the geometric mean is less than or equal to the arithmetic mean. It follows that

$$|f(x+iy)|^p \leqq (1/\pi) e^{pky} \int_{-\infty}^{\infty} \frac{y}{(x-t)^2 + y^2} |f(t)|^p\, dt.$$

The above integral converges uniformly for x in the closed interval $[-T, T]$ for, if

$$\alpha(t) = \int_0^t |f(x)|^p\, dx, \qquad \beta(t) = \frac{y}{(x-t)^2 + y^2},$$

and $M^p\{f(x)\} = A$, then given $\epsilon > 0$, there exists a t_0 such that for all $t > t_0$,

$$\alpha(t) \leq \int_{-t}^{t} | f(x) |^p \, dx \leqq 2(A + \epsilon) | t | .$$

Now

$$\int_0^\infty \frac{y}{(x-t)^2 + y^2} | f(t) |^p \, dt = \int_0^\infty \beta(t) \, d\alpha(t).$$

Integrating by parts, it is seen that

$$\left| \int_0^\infty \frac{y}{(x-t)^2 + y^2} | f(t) |^p \, dt \right| = \left| \int_0^\infty \alpha(t) \, d\beta(t) \right|$$
$$< \left| \int_0^{t_0} \alpha(t) \, d\beta(t) \right| + \left| \int_{t_0}^\infty \frac{2y(A+\epsilon)t(x-t)}{\{(x-t)^2 + y^2\}^2} \, dt \right| < \infty,$$

uniformly for x in $[-T, T]$. The integral over $[-\infty, 0]$ is treated similarly.

The expression for $| f(x+iy) |^p$ can then be integrated with respect to x over the interval $[-T, T]$, and the order of integration reversed, i. e.,

$$/2T) \int_{-T}^{T} | f(x+iy) |^p \, dx \leq e^{pky} (1/\pi) \int_{-\infty}^{\infty} \{ (1/2T) \int_{-T}^{T} \frac{y}{(x-t)^2 + y^2} | f(t) |^p \, dx \} d$$

Now let

$$\omega_T(t) = \int_{-T}^{T} \frac{y}{(x-t)^2 + y^2} dx, \qquad \gamma(t) = \int_{-t}^{t} | f(x) |^p \, dx.$$

Noting that $\omega_T(-t) = \omega_T(t)$, it follows that

$$(1/2T) \int_{-T}^{T} | f(x+iy) |^p \, dx \leq e^{pky} (1/2\pi T) \int_0^\infty \omega_T(t) \, d\gamma(t).$$

Integrating by parts (recalling that, as in the case of $\alpha(t)$, $\gamma(t) = O(t)$ as $t \to \infty$),

$$(1/2T) \int_{-T}^{T} | f(x+iy) |^p \, dx = - e^{pky} (1/2\pi T) \int_0^\infty \gamma(t) \, d\omega_T(t)$$
$$< - e^{pky} (1/2\pi T) \{ \int_0^{t_0} \gamma(t) \, d\omega_T(t) + (A+\epsilon) \int_{t_0}^\infty t \, d\omega_T(t) \}.$$

Evaluating the last two integrals, it is seen that

$$\overline{\lim_{T\to\infty}} \, (1/2\pi T) \int_0^{t_0} \gamma(t) \, d\omega_T(t) = 0,$$

$$\overline{\lim_{T\to\infty}} \, (1/2\pi T) \int_{t_0}^\infty t \, d\omega_T(t) = -1,$$

hence

$$\overline{\lim_{T\to\infty}} \, (1/2T) \int_{-T}^{T} | f(x+iy) |^p \, dx \leq (A+\epsilon) e^{pky}.$$

Since ϵ is arbitrary, the theorem is proved.

3. The mean of $f'(x)$. It will now be shown that the boundedness of $M^p\{f(x)\}$ implies the boundedness of $M^p\{f'(x)\}$. First, a rather crude inequality will be obtained in the case where p is greater than zero. When p is greater than one, a refinement of the crude result can be obtained, as will be seen.

LEMMA 1. *If a is a real number, then*

$$M^p\{f(x+a+iy)\} = M^p\{f(x+iy)\}, \qquad (p>0).$$

Proof. Assume $a>0$. (The proof is the same for $a<0$.) Then,

$$\frac{T-a}{T}\cdot\frac{1}{2(T-a)}\int_{-(T-a)}^{(T-a)}|f(x+iy)|^p\,dx \leqq (1/2T)\int_{-T}^{T}|f(x+a+iy)|^p\,dx$$
$$\leqq \frac{T+a}{T}\cdot\frac{1}{2(T+a)}\int_{-(T+a)}^{(T+a)}|f(x+iy)|^p\,dx.$$

Since $(T-a)/T$ and $(T+a)/T$ approach 1 as T approaches infinity, the result follows.

THEOREM 3. *If $f(z)$ is of exponential type k, then*

$$M^p\{f'(x)\} \leqq \frac{(p+2)2^{p+2}}{\pi k p\delta^{p+1}}(e^{k\delta p}-1)M^p\{f(x)\}, \qquad (p>0),$$

where δ is an arbitrary positive number.

Proof. It is known [24; p. 127] that if $f(z)$ is regular in a square with corners $(x+\delta\pm i\delta)$, $(x-\delta\pm i\delta)$, then

$$|f'(x)|^p \leq A\int_{-\delta}^{\delta}\int_{-\delta}^{\delta}|f(x+s+it)|^p\,dsdt,$$

where δ is an arbitrary positive number and $A=2^p(p+2)/(\pi\delta^{p+2})$. Hence,

$$(1/2T)\int_{-T}^{T}|f'(x)|^p\,dx \leq A\int_{-\delta}^{\delta}\int_{-\delta}^{\delta}\{(1/2T)\int_{-T}^{T}|f(x+u+iv)|^p\,dx\}dudv.$$

By the preceding lemma and Theorem 2, if u and v are fixed,

$$M^p\{f(x+u+iv)\} = M^p\{f(x+iv)\} \leq e^{pk|v|}M^p\{f(x)\}.$$

It follows that, given $\epsilon>0$, there exists a $T_0(u,v,\epsilon)$, such that

$$(1/2T)\int_{-T}^{T}|f(x+u+iv)|^p\,dx \leq e^{pk|v|}M^p\{f(x)\}+\epsilon$$

for all $T>T_0$. Such a T_0 exists for each u and v with $|u|\leq\delta$, $|v|\leq\delta$. Since this set of values for u and v is closed, $T_0(u,v,\epsilon)$, as a function of

u and v, must be bounded above for u, v in the closed interval $[-\delta, \delta]$. Let T_1 be an upper bound of $T_0(u, v)$ for u, v on $[-\delta, \delta]$. The last inequality then holds for all $T > T_1$. Hence, for $T > T_1$,

$$(1/2T) \int_{-T}^{T} | f'(x) |^p \, dx \leq A \int_{-\delta}^{\delta} \int_{-\delta}^{\delta} [e^{pk|v|} M^p\{f(x)\} + \epsilon] du dv$$
$$= A 4\delta [(\frac{e^{pk\delta}}{pk} - 1) M^p\{f(x)\} + \epsilon].$$

Letting T approach infinity first and then letting ϵ approach zero, the result follows.

It should be noted that by repeated application of this theorem a bound for $M^p\{f^{(n)}(x)\}$ is obtained.

It will now be shown that when p is greater than one, the constant in Theorem 3 can be replaced by k^p.

LEMMA 2. *If*

$$g_\epsilon(z) = \frac{\sin \epsilon z}{\epsilon z} f(z),$$

then

$$\lim_{\epsilon \to 0} g'_\epsilon(z) = f'(z).$$

Proof.

$$g'_\epsilon(z) = \frac{\sin \epsilon z}{\epsilon z} f'(z) + \frac{f(z)}{\epsilon^2 z^2} (\epsilon^2 \cos \epsilon z - \sin \epsilon z).$$

If $\cos \epsilon z$ and $\sin \epsilon z$ in the last term are expanded into their Maclaurin's series, it is seen that the last term approaches zero with ϵ.

LEMMA 3. *If $f(z)$ is of exponential type $k < \pi$, and*

$$| f(x) | \leq B < \infty,$$

then

$$f'(x) = (1/\pi) \sum_{n=-\infty}^{\infty} \frac{(-1)^{n+1} f(x + n + \frac{1}{2})}{(n - \frac{1}{2})^2}.$$

Proof. If ϵ is chosen less than $(\pi - k)$, then the function

$$g_\epsilon(z) = \frac{\sin \epsilon (z + x_0 - \frac{1}{2})}{\epsilon (z + x_0 - \frac{1}{2})} f(z + x_0 - \tfrac{1}{2})$$

is an entire function of exponential type less than π and satisfies

$$\sum_{n=-\infty}^{\infty} n^{-1} g_\epsilon(n) < \infty.$$

By Valiron's interpolation formula [see 34], or from Pólya and Szegö [27; vol. I, p. 16, no. 165],

$$\frac{g'_\epsilon(z)}{\sin \pi z} - \frac{\pi g_\epsilon(z)}{\sin \pi z \tan \pi z} = -(1/\pi) \sum_{n=-\infty}^{\infty} \frac{(-1)^n g_\epsilon(n)}{(z-n)^2}.$$

Letting z approach $\frac{1}{2}$,

$$g'_\epsilon(\tfrac{1}{2}) = -(1/\pi) \sum_{n=-\infty}^{\infty} \frac{(-1)^n g_\epsilon(n)}{(n-\frac{1}{2})^2}.$$

Since $|g_\epsilon(x)|$ is bounded for all ϵ, this series converges uniformly in ϵ. Letting ϵ approach zero, and using the result of the previous lemma,

$$f'(x_0) = -(1/\pi) \sum_{n=-\infty}^{\infty} \frac{(-1)^n f(n + x_0 - \frac{1}{2})}{(n-\frac{1}{2})^2}.$$

This is true for each x_0; hence the lemma is proved.

Lemma 4. *If $F(n,T)$ is positive or zero for all integers n and all real T, and if $\sum_{n=-\infty}^{\infty} F(n,T)$ converges uniformly in T, $-\infty < T < \infty$, then*

$$\overline{\lim_{T\to\infty}} \sum_{n=-\infty}^{\infty} F(n,T) \leq \sum_{n=-\infty}^{\infty} \overline{\lim_{T\to\infty}} F(n,T).$$

Proof. By hypothesis, given $\epsilon > 0$, there exists a positive integer N (depending only on ϵ) such that

$$\sum_{n=-\infty}^{\infty} F(n,T) \leq \sum_{n=-N}^{N} F(n,T) + \epsilon.$$

Hence,

$$\overline{\lim_{T\to\infty}} \sum_{n=-\infty}^{\infty} F(n,T) \leq \sum_{n=-N}^{N} \overline{\lim_{T\to\infty}} F(n,T) + \epsilon \leq \sum_{n=-\infty}^{\infty} \overline{\lim_{T\to\infty}} F(n,T) + \epsilon.$$

The result then follows on letting ϵ approach zero.

Theorem 4. *If $f(z)$ is of exponential type k, then for $p > 1$,*

$$M^p\{f'(x)\} \leq k^p M^p\{f(x)\}.$$

Proof.

Case I: $k < \pi$.

It is assumed that $M^p\{f(x)\}$ is bounded, otherwise there is nothing to prove. Then, since $p > 1$, $f(x) = O(|x|)$, by Theorem 1.

Choose $\epsilon > 0$ such that $k + \epsilon < \pi$, and consider

$$g_\epsilon(z) = \frac{f(z) \sin \epsilon z}{\epsilon z}.$$

The function $g_\epsilon(z)$ is of type $k + \epsilon < \pi$, and since $f(x) = O(|x|)$, $g_\epsilon(x)$ is bounded. By Lemma 3,

$$g'_\epsilon(x) = (1/\pi) \sum_{n=-\infty}^{\infty} \frac{(-1)^{n+1} g_\epsilon(x+n-\frac{1}{2})}{(n-\frac{1}{2})^2}.$$

Applying Minkowski's inequality,

$$\{(1/2T) \int_{-T}^{T} | g'_\epsilon(x) |^p \, dx\}^{1/p} \leq (1/\pi) \sum_{n=-\infty}^{\infty} \{(1/2T) \int_{-T}^{T} \left| \frac{g_\epsilon(x+n-\frac{1}{2})}{(n-\frac{1}{2})^2} \right|^p dx\}^{1/p}$$

Since $| g_\epsilon(x) |$ is less than or equal to $| f(x) |$ for all ϵ, $| g_\epsilon(x+n-\frac{1}{2}) |$ may be replaced by $| f(x+n-\frac{1}{2}) |$ in the above inequality. The right-hand side then does not depend on ϵ; hence, letting ϵ approach zero and using Lemma 2,

$$(1) \ \{(1/2T) \int_{-T}^{T} | f'(x) |^p \, dx\}^{1/p} \leq (1/\pi) \sum_{n=-\infty}^{\infty} \{(1/2T) \int_{-T}^{T} \left| \frac{f(x+n-\frac{1}{2})}{(n-\frac{1}{2})^2} \right|^p dx\}^{1/p}$$

At this point it is desirable to take the limit superior inside the above sum. In order to do this it will be shown that the sum satisfies the hypothesis of Lemma 4.

Let the sum be broken up into three sums:

$$S_1 + S_2 + S_3 = \sum_{-\infty}^{-1} + \sum_{0}^{1} + \sum_{2}^{\infty}.$$

By hypothesis, there exists an $A < \infty$ such that

$$(1/2T) \int_{-T}^{T} | f(x) |^p \, dx < A$$

for all T.

If n is less than zero and T is greater than three, it follows that $(\frac{1}{2}-n)T$ is greater than $(T+\frac{1}{2}-n)$; hence for $T > 3$,

$$S_1 < \sum_{-\infty}^{1} \frac{1}{| n-\frac{1}{2} |^{2-1/p}} \left\{ \frac{1}{2(T+\frac{1}{2}-n)} \int_{-(T+\frac{1}{2}-n)}^{T+\frac{1}{2}-n} | f(x) |^p \, dx \right\}^{1/p}$$

$$< \sum_{-\infty}^{-1} \frac{A^{1/p}}{| n-\frac{1}{2} |^{2-1/p}}.$$

Since p is greater than one, this converges and so S_1 converges uniformly in T, $(T > 3)$. A similar argument shows that S_3 converges uniformly in T, $(T > 3)$.

The sum in (1) then satisfies Lemma 4 and the upper limit may be taken inside the sum, giving

$$[M^p\{f'(x)\}]^{1/p} \leq (1/\pi) [M^p\{f(x)\}]^{1/p} \sum_{-\infty}^{\infty} 1/(n-\tfrac{1}{2})^2.$$

Since the last sum is equal to π^2, it follows that

(2) $$M^p\{f'(x)\} \leq \pi^p M^p\{f(x)\}.$$

CASE II: general k.

Let ϵ be an arbitrary positive number and consider

$$h(z) = f\left(\frac{(1-\epsilon)\pi z}{k}\right).$$

The function $h(z)$ is of type less than π. Applying (2) to $h(z)$, it follows that

$$(1-\epsilon)^p(\pi/k)^p M^p\{f'(x)\} \leq \pi^p M^p\{f(x)\}.$$

The general result then follows on letting ϵ approach zero.

By considering $\sin kz$ (which is of type k) it is seen that the constant k^p is "best possible."

4. Functions of minimal type. A function of exponential type zero is said to be of minimal type. If k is made to approach zero in Theorem 2 and Theorem 4, it is seen that $M^p\{f(x+iy)\} = M^p\{f(x)\}$ and $M^p\{f'(x)\} = 0$, the latter being true for $p > 1$. This would seem to indicate that $f(z)$ is a constant. It will now be shown that this is indeed true.

LEMMA 5. *If $f(x)$ is a continuous function such that $|f(x)|$ becomes and remains larger than any assigned constant as $|x|$ approaches infinity, then $M^p\{f(x)\} = \infty$, $(p > 0)$.*

Proof. Given B, there exists $X_0 > 0$, such that $|f(x)|$ is greater than B for all $x > X_0$. Hence,

$$M^p\{f(x)\} > \lim_{T\to\infty} (1/2T) \int_{-X_0}^{X_0} |f(x)|^p\, dx + \lim_{T\to\infty} (B^p/T)(T - X_0) = B^p.$$

Since B is arbitrary, the lemma is proved.

THEOREM 5. *If $f(z)$ is of minimal type and $M^p\{f(x)\} < \infty$, $(p > 0)$, then $f(z)$ is identically a constant.*

Proof. Let $r = [1/p] + 1$, where $[1/p]$ denotes the integral part of $1/p$, and consider the function

$$g(z) = \{f(z) - P_r(z)\}z^{-r},$$

where $P_r(z)$ denotes the first r terms of the Maclaurin's series of $f(z)$.

The function $g(z)$ is entire and of minimal type. Since, by Theorem 1,

$$f(x) = O(|x|^{1/p}) = o(|x|^r),$$

it follows that $g(x)$ is bounded. By [25] or [20; p. 127], $g(z)$ is identically a constant, and hence $f(z)$ is a polynomial of degree at most r. A polynomial of degree greater than zero, however, satisfies the hypothesis of the preceding lemma. By the hypothesis of this theorem; this is impossible, so the degree of the polynomial must be zero.

5. Continuity of the function $\phi(y) = M^p\{f(x+iy)\}$, $(p \geq 1)$.

THEOREM 6. *If $f(z)$ is of exponential type k, and if $M^p\{f(x)\}$ is finite, then the function $\phi(y) = M^p\{f(x+iy)\}$, $p \geq 1$, is a continuous function of y.*

Proof. Assume that $y < y'$. (The proof is the same for $y > y'$.) From the definition of $\phi(y)$ and the usual inequalities on upper limits,

$$|\phi(y) - \phi(y')| \leq \overline{\lim_{T\to\infty}}\,(1/2T)\int_{-T}^{T}\Big|\,|f(x+iy)|^p - |f(x+iy)|^p\Big|\,dx.$$

Now,

$$\begin{aligned}\Big|\,|f(x+iy)|^p - |f(x+iy')|^p\Big| &= \left|\int_y^{y'}\frac{\partial}{\partial u}|f(x+iu)|^p\,du\right| \\ &\leq \int_y^{y'}\left|\frac{\partial}{\partial u}|f(x+iu)|^p\,du\right| \\ &\leq \int_y^{y'} p\,|f(x+iu)|^{p-1}\,|f'(x+iu)|.\end{aligned}$$

Using Hölder's inequality when $p > 1$ (and trivially for $p = 1$),

$$\begin{aligned}|\phi(y) - \phi(y')| \leq p\,\overline{\lim_{T\to\infty}}\int_y^{y'}\Big\{\Big[(1/2T)\int_{-T}^{T}|f'(x+iu)|^p\,dx\Big]^{1/p} \\ \cdot\Big[(1/2T)\int_{-T}^{T}|f(x+iu)|^p\Big]^{1-1/p}\Big\}du.\end{aligned}$$

By Theorem 2, $M^p\{f(x+iu)\}$ is bounded, hence there exists a $B_1(u)$ such that for all T,

$$(1/2T)\int_{-T}^{T}|f(x+iu)|^p\,dx < B_1(u).$$

By Theorem 3, $M^p\{f'(x)\}$ is bounded. Furthermore, if $f(z)$ is of exponential type k, so is $f'(z)$ [35; p. 41]; hence Theorem 2 can be applied to $f'(z)$. Then there exists a $B_2(u)$ such that for all T,

$$(1/2T)\int_{-T}^{T}|f'(x+iu)|^p\,dx < B_2(u).$$

It follows that

$$|\phi(y)-\phi(y')| \leq p(y-y') \max_{y\leq u\leq y'} [\{B_1(u)\}^{1/p}\{B_2(u)\}^{1-1/p}],$$

which proves the continuity of $\phi(y)$.

6. The mean of $f(z)$ at the integers. It will now be shown that the boundedness of $M^p\{f(x)\}$ implies the boundedness of $M_i{}^p\{f(x)\}$ and that the converse is true if the type of $f(z)$ is less than π. The device which will be used is essentially the same as that used by Plancherel and Pólya [24] in obtaining analogous theorems for functions of class L^p on the axis.

LEMMA 6. *Let n be a positive integer and let c be a positive constant. If $h(x)$ is a positive and continuous function, then*

$$\overline{\lim_{T\to\infty}}\,(1/2T)\int_{-T}^{T} h(x)\,dx = \overline{\lim_{n\to\infty}}\,(1/2nc)\int_{-nc}^{nc} h(x)\,dx.$$

Proof. Let $[T]$ denote the integral part of T. The result then follows from the inequalities:

$$\frac{[T]}{T}\cdot\frac{1}{2[T]c}\int_{-[T]c}^{[T]c} h(x)\,dx \leq (1/2Tc)\int_{-Tc}^{Tc} h(x)\,dx \leq \frac{[T+1]}{T}\cdot\frac{1}{2[T+1]c}\int_{-[T+1]c}^{[T+1]c} h(x)\,dx.$$

THEOREM 7. *If $f(z)$ is of exponential type k, then*

$$M_i{}^p\{f(x)\} = \frac{8(e^{\frac{1}{2}kp}-1)}{\pi kp}\,M^p\{f(x)\}.$$

Proof. Plancherel and Pólya [24; p. 125] have shown that if $f(z)$ is regular in the square with corners $(j+\frac{1}{2}\pm\frac{1}{2}i)$, $(j-\frac{1}{2}\pm\frac{1}{2}i)$, then

$$|f(j)|^p \leq (4/\pi)\int_{-1/2}^{1/2}\int_{-1/2}^{1/2}|f(j+x+iy)|^p\,dxdy = (4/\pi)\int_{-1/2}^{1/2}\int_{-1/2+j}^{1/2+j}|f(x+iy)|^p\,dxdy;$$

hence,

$$\frac{1}{2n+1}\sum_{j=-n}^{n}|f(j)|^p \leq \frac{4}{\pi(2n+1)}\int_{-1/2}^{1/2}\left\{\int_{-1/2-n}^{1/2-n}+\cdots+\int_{-1/2+n}^{1/2+n}|f(x+iy)|^p\,dx\right\}dy$$

$$= (4/\pi)\int_{-1/2}^{1/2}\left\{\frac{1}{2(n+\frac{1}{2})}\int_{-(n+1/2)}^{n+1/2}|f(x+iy)|^p\,dx\right\}dy.$$

Applying Lemma 6 and Theorem 2, it follows that, given $\epsilon>0$, there exists an N such that for all $n>N$ and all y with $-\frac{1}{2}\leq y\leq\frac{1}{2}$,

$$\frac{1}{2n+1}\int_{-(n+1/2)}^{n+1/2}|f(x+iy)|^p\,dx < e^{pk|y|}M^p\{f(x)\}+\epsilon.$$

Consequently,

$$\frac{1}{2n+1}\sum_{j=-n}^{n}|f(j)|^p \leq (4/\pi)M^p\{f(x)\}\int_{-1/2}^{1/2} e^{pk|y|}dy + 4\epsilon/\pi.$$

The result follows on evaluating the integral and letting n approach infinity first and then ϵ approach zero.

The fact that the boundedness of $M^p\{f(x)\}$ implies the boundedness of $M_i^p\{f(x)\}$ is not surprising. The converse is certainly a more interesting speculation. One has only to consider $f(z) = z \sin \pi z$ (which is of type π) to realize that the converse is not true in general. It will be shown, however, that if the type of $f(z)$ is less than π, the converse does indeed hold.

LEMMA 7. *Let p be positive and $M_i^p\{f(x)\}$ finite. If n is an integer, then*

$$f(n) = O(|n|^{1/p}), \text{ as } |n| \to \infty.$$

Proof. The proof follows almost immediately from the definition of $M_i^p\{f(x)\}$.

LEMMA 8. *If m is a fixed integer,*

$$M_i^p\{f(x+m)\} = M_i^p\{f(x)\}, \qquad p > 0.$$

Proof. The proof is essentially the same as for Lemma 1, integrals being replaced by sums.

THEOREM 8. *If $f(z)$ is of exponential type $k < \pi$, then*

$$M^p\{f(x)\} \leq BM_i^p\{f(x)\}, \qquad (p > 0),$$

where B is a constant depending on p and k only.

Proof. Assume that $M_i^p\{f(x)\}$ is bounded; otherwise there is nothing to prove.

Let $a = 2[1/p]$, where $[1/p]$ denotes the integral part of $1/p$. Let $\delta = \frac{1}{2}(\pi - k)/(a+3)$, and consider the function

$$g(z) = z^{-a} f(z+m) \sin^{a+3} \delta z, \qquad (m \text{ fixed}).$$

The function $g(z)$ is entire and of type less than or equal to $k + (a+3)\delta$. By definition of δ, the type of $g(z)$ is less than π.

It will now be shown that $g(z)$ satisfies

$$\lim_{z \to \infty} |g(z)|\,|z|^{-2} e^{-\pi|z|} = 0, \tag{3}$$

and

$$\sum_{n=-\infty}^{\infty}{}' |g(n) n^{-3}| < \infty. \tag{4}$$

(The primed summation indicates that $n = 0$ is to be omitted.)

The first relationship follows immediately from the fact that $g(z)$ is of type less than π.

By definition of $g(z)$,

$$\sum_{n=-\infty}^{\infty}{}' | n^{-3}g(n) | \leq \sum_{n=-\infty}^{\infty}{}' | f(n+m)n^{-(a+3)} | .$$

$$\leq \sum_{n=-\infty}^{-m-1}{}' + \sum_{n=m}{}' + \sum_{n=m+1}^{\infty}{}' = S_1 + S_2 + S_3.$$

Now,

$$S_1 = \sum_{n=-\infty}^{-m-1} \left| \frac{f(n+m)}{(n+m)^{\frac{1}{2}a+1}} \right| \cdot \left| \frac{(n+m)^{\frac{1}{2}a+1}}{n^{a+3}} \right| .$$

Since by Lemma 7, $f(n+m) = O(\{n+m\}^{1/p})$, it follows from the definition of "a" that $f(n+m)/(n+m)^{\frac{1}{2}a+1}$, in S_1, is bounded, say by A. Then, using the binomial expansion of $(n+m)^{\frac{1}{2}a+1}$,

$$S_1 \leq A \sum_{n=-\infty}^{-m-1}{}' \left(\frac{1}{n^2} + \frac{(\frac{1}{2}a+1)m}{n^3} + \cdots + \frac{m^{\frac{1}{2}a+1}}{n^{a+3}} \right) < \infty.$$

In a similar way, S_3 is seen to converge; hence (4) is satisfied.

Since $g(z)$ satisfies (3), (4), Valiron's interpolation formula [34] may be applied, i. e.

$$g(z) = \frac{\sin \pi z}{\pi} \sum_{n=-\infty}^{\infty}{}' \frac{(-1)^n z^2 g(n)}{n^2(z-n)},$$

or, by definition of $g(z)$,

$$f(z+m) = \frac{z^{a+2}\sin \pi z}{\pi(\sin \delta z)^{a+3}} \sum_{n=-\infty}^{\infty}{}' \frac{(-1)^n f(n+m) \cdot (\sin \delta n)^{a+3}}{(z-n)n^{a+2}}. \tag{5}$$

Let F_m be the maximum of $| f(x) |$ on the interval, $m - \frac{1}{2} \leq x \leq m + \frac{1}{2}$. Then by the maximum modulus principle, F_m is less than or equal to the maximum of $|f(x+m)|$ on the square S, with corners: $\pm \frac{1}{2}(1+i)$, $\pm (1-i)$. Let

$$\max_{z \text{ on } S} \left(\frac{z^{a+2}\sin \pi z}{\pi(\sin \delta z)^{a+3}} \right) = K.$$

On S, $| z - n |$ is greater than $| n | - \frac{1}{2}$, hence using the interpolation formula (5),

$$F_m \leq K \sum_{n=-\infty}^{\infty}{}' Q(m, n), \tag{6}$$

where

$$Q(m,n) = \frac{| f(n+m)(\sin \delta n)^{a+3} |}{(| n | - \frac{1}{2}) | n |^{a+2}}.$$

Now,

$$M^p\{f(x)\} = \overline{\lim_{r\to\infty}} \frac{1}{2(r+\frac{1}{2})} \left\{ \int_{-r-\frac{1}{2}}^{-r+\frac{1}{2}} + \cdots + \int_{r-\frac{1}{2}}^{r+\frac{1}{2}} |f(x)|^p \, dx \right\},$$

where r can approach infinity through integral values, by Lemma 6. Hence, from the definition of F_m,

$$(7) \qquad M^p\{f(x)\} \leq \overline{\lim_{r\to\infty}} \frac{1}{2r+1} \sum_{m=-r}^{r} (F_m)^p.$$

Two cases are now considered.

CASE I: $p > 1$.

In the same way as the summation in (4) was seen to converge, it is seen that $\Sigma' Q(m,n)$ converges uniformly for m in the interval, $-r \leq m \leq r$. Using (6) and Minkowski's inequality,

$$\{\sum_{m=-r}^{r} (F_m)^p\}^{1/p} \leq K\{\sum_{m=-r}^{r} (\sum_{n=-\infty}^{\infty}{}' Q(m,n))^p\}^{1/p}$$

$$\leq K \sum_{n=-\infty}^{\infty}{}' \{\sum_{m=-r}^{r} Q^p(m,n)\}^{1/p}.$$

It follows that

$$(8) \qquad [M^p\{f(x)\}]^{1/p} \leq K \overline{\lim_{r\to\infty}} \sum_{n=-\infty}^{\infty}{}' \{\frac{1}{2r+1} \sum_{m=-r}^{r} Q^p(m,n)\}^{1/p}.$$

Next, it will be shown that the upper limit can be taken inside the first sum in relation (8).

Let

$$\sum_{n=-\infty}^{\infty}{}' = \sum_{n=-\infty}^{n=-2} + \sum_{n=-1}^{1}{}' + \sum_{n=2}^{\infty} = S_1 + S_2 + S_3.$$

From the definition of $Q(m,n)$,

$$\left\{ \frac{1}{2r+1} \sum_{m=-r}^{r} Q^p(m,n) \right\}^{1/p} = \frac{|\sin^3 \delta n|}{(|n|-\frac{1}{2})|n|^{2-1/p}} \left\{ \frac{1}{(2r+1)|n|} \sum_{m=-r+n}^{r+n} |f(m)|^p \right\}^{1/p}.$$

In S_1,

$$(2r+1)|n| > 2(r-n)$$

if $r > 1$, hence,

$$S_1 \leq \sum_{n=-\infty}^{\infty}{}' \frac{1}{(|n|-\frac{1}{2})|n|^{2-1/p}} \left\{ \frac{1}{2(r-n)} \sum_{m=-(r-n)}^{r-n} |f(m)|^p \right\}^{1/p}.$$

Since $M_i^p\{f(x)\}$ is finite, the expression in the brackets is bounded, and since p is greater than one, the sum converges uniformly in r.

In a similar way, it can be shown that S_3 converges uniformly in r; hence, by Lemma 4, the upper limit can be taken inside the sum in (8) giving

$$[M^p\{f(x)\}]^{1/p} \leq K \sum_{n=-\infty}^{\infty}{}' \frac{\sin^3 \delta n}{(|n|-\frac{1}{2})|n|^2} \{\overline{\lim_{r\to\infty}} \frac{1}{2r+1} \sum_{m=-r}^{r} |f(n+m)|^p\}^{1/p}$$

$$= K[M_i^p\{f(x)\}]^{1/p} \sum_{n=-\infty}^{\infty}{}' \frac{\sin^3 \delta n}{(|n|-\frac{1}{2})|n|^2},$$

the last step being a consequence of Lemma 8. This sum converges; hence the result is proved for $p > 1$, with

$$B = K^p \left\{ \sum_{n=-\infty}^{\infty}{}' \frac{\sin^3 \delta n}{(|n|-\frac{1}{2})|n|^2} \right\}^p.$$

CASE II: $p \leq 1$.

Returning to (6), and using Jensen's inequality,

$$\sum_{m=-r}^{r} (F_m)^p \leq K^p \sum_{m=-r}^{r} \{ \sum_{n=-\infty}^{\infty}{}' Q(m,n)\}^p \leq K^p \sum_{m=-r}^{r} \{ \sum_{n=-\infty}^{\infty}{}' Q^p(m,n)\}.$$

Now,

$$Q^p(m,n) \leq \frac{|f(n+m)\sin \delta n)^{a+3}|^p}{(|n|-\frac{1}{2})^p|n|^{(a+2)p}} \leq \frac{|f(n+m)|^p}{(|n|-\frac{1}{2})^p|n|^2}$$

by definition of a. It follows, then, that

$$\sum_{n=-\infty}^{\infty}{}' Q^p(m,n) \leq \sum_{n=-\infty}^{\infty}{}' \frac{|f(n+m)|^p}{|n+m|} \cdot \left| \frac{n+m}{(n-\frac{1}{2})^p n^2} \right|.$$

Since $f(n+m) = O(\{n+m\}^{1/p})$ by Lemma 7, the sum is seen to converge uniformly for m in the interval $-r \leq m \leq r$. The order of summation can then be changed, and so by (7)

$$M^p\{f(x)\} \leq K^p \overline{\lim_{r\to\infty}} \sum_{n=-\infty}^{\infty} \{\frac{1}{2r+1} \sum_{m=-r}^{r} Q^p(m,n)\}. \tag{9}$$

Since

$$p(a+2) > 2,$$

it follows that

$$\sum_{n=-\infty}^{\infty}{}' \{\frac{1}{2r+1} \sum_{m=-r}^{r} Q^p(m,r)\} \leq \sum_{n=-\infty}^{\infty}{}' \frac{|\sin \delta n|^{(a+3)p}}{|n|(|n|-\frac{1}{2})^p} \left\{ \frac{1}{2r|n|} \sum_{m=-r}^{r} |f(n+m)|^p \right\}.$$

By an argument similar to the one used in Case I, the sum can be split into three sums so that the term in brackets is bounded in the infinite sums. Consequently the sum is seen to converge uniformly in r. The upper limit in (9) can then be taken inside the infinite sum. The result then follows as in Case I with

$$B = K^p \sum_{n=-\infty}^{\infty}{}' \left| \frac{(\sin \delta n)^{a+3}}{n^{a+2}(|n|-\frac{1}{2})} \right|^p \leq K^p \sum_{n=-\infty}^{\infty} \frac{1}{|n|^2(|n|-\frac{1}{2})^p} < \infty.$$

7. One-sided means.

DEFINITION. The function $f(z)$ is said to be of exponential type k in $R(z) \geqq 0$ if $f(z)$ is regular in $R(z) \geqq 0$ and

$$M(r) = O(e^{(k+\epsilon)r})$$

for every $\epsilon > 0$ and no $\epsilon < 0$, $M(r)$ being the maximum of $|f(z)|$ on the semicircle $|z| = r$, $R(z) \geqq 0$.

DEFINITION. The one-sided p-th mean of $f(z)$ on a horizontal half-line, denoted by $N^p\{f(x+iy)\}$, is defined as follows:

$$N^p\{f(x+iy)\} = \overline{\lim_{n\to\infty}} (1/T) \int_0^T |f(x+iy)|^p \, dx.$$

The one-sided p-th mean of $f(z)$ at the integers, denoted by $N_i^p\{f(x)\}$, is defined similarly:

$$N_i^p\{f(x)\} = \overline{\lim_{n\to\infty}} (n+1)^{-1} \sum_{i=0}^{n} |f(i)|^p.$$

It will be shown that the main theorems for two-sided means hold also for one-sided means. A theorem of Macintyre [21] will be useful and the theorem is stated here as a lemma. It was not stated in this form by Macintyre but an inspection of his proof will bear out the validity of the lemma.

LEMMA 9. *Let $f(z)$ be of exponential type $k < \pi$ in $R(z) \geqq 0$, and assume that*

$$\overline{\lim_{r\to\infty}} r^{-1} \log |f(r)| \leqq 0.$$

Then $f(z)$ can be written as the sum of two functions $f_1(z)$, $f_2(z)$, where $f_1(z)$ is entire and of type k. If y is fixed,

$$f_1(x+iy) = O(1/x) \qquad \text{as } x \to -\infty$$

$$f_2(x+iy) = O(1/x) \qquad \text{as } x \to +\infty.$$

8. One-sided means on a half-line.

LEMMA 10. *Let*

$$f(z) = f_1(z) + f_2(z),$$

and let

$$N^p\{f_2(x+iy)\} = 0, \qquad (p > 0).$$

Then

$$N^p\{f(x+iy)\} = N^p\{f_1(x+iy)\}.$$

Proof. The proof follows immediately by applying Minkowski's inequality for $p \geqq 1$, or Jensen's inequality for $p \leqq 1$, to the relations:

$$f(x+iy) = f_1(x+iy) + f_2(x+iy),$$

$$f_1(x+iy) = f(x+iy) - f_2(x+iy).$$

LEMMA 11. *If $\phi(x)$ is continuous for $x \geqq 0$, and $\phi(x) \to 0$ as $x \to \infty$, then*

$$N^p\{\phi(x)\} = 0, \qquad p > 0.$$

Proof. Given $\epsilon > 0$, there exists X_0 such that for all $x > X_0$,

$$|\phi(x)| < \epsilon.$$

Hence, if $T > X_0$,

$$(1/T)\int_0^T |\phi(x)|^p\,dx = (1/T)\int_0^{X_0} |\phi(x)|^p + (1/T)\int_{X_0}^T |\phi(x)|^p\,dx$$
$$\leq A^p X_0/T + \epsilon^p (T - X_0)/T,$$

where A is the maximum of $\phi(x)$ on $0 \leq x \leq X_0$. It follows that

$$N^p\{\phi(x)\} \leq \epsilon^p.$$

Letting ϵ approach zero, the lemma is proved.

LEMMA 12. *If $\phi(x)$ is continuous and $\phi(x) \to 0$ as $x \to -\infty$, then*

$$2M^p\{\phi(x)\} = N^p\{\phi(x)\}.$$

Proof.

$$2(1/2T)\int_{-T}^T |\phi(x)|^p\,dx = (1/T)\int_0^T |\phi(-x)|^p\,dx + (1/T)\int_0^T |\phi(x)|^p\,dx.$$

From the previous lemma,

$$\overline{\lim_{T\to\infty}}\,(1/T)\int_0^T |\phi(-x)|^p\,dx = 0,$$

hence the upper limit equals the lower limit. Using the well known inequalities on upper and lower limits, it is easily seen that

$$N^p\{\phi(x)\} \leq 2M^p\{\phi(x)\} \leq N^p\{\phi(x)\},$$

and the lemma is proved.

THEOREM 9. *If $f(z)$ is of exponential type k in $R(z) \geqq 0$, then*

$$N^p\{f(x+iy)\} \leq e^{pk|y|}N^p\{f(x)\}, \qquad (p > 0).$$

Proof. Assume that $k < \pi$. (Otherwise consider $f(z/k)$.)

By the first part of Theorem 1, $f(x) = O(x^{1/p})$; hence

$$\overline{\lim_{r\to\infty}}\, r^{-1} \log |f(r)| \leq 0.$$

Lemmas 9, 10, 12, and Theorem 2 can then be applied, giving

$$N^p\{f(x+iy)\} = N^p\{f_1(x+iy)\} = 2M^p\{f_1(x+iy)\}$$
$$\leq e^{pk|y|}\, 2M^p\{f_1(x)\} = e^{pk|y|} N^p\{f(x)\}.$$

9. The one-sided mean of $f'(x)$.

LEMMA 13. *Let $f(z)$ be of exponential type k in $R(z) \geq 0$, and let $f(x)$ be bounded for all $x \geq 0$; then $f'(x)$ is bounded for all $x \geq 0$.*

Proof. It is known [13; p. 143] that under the above hypothesis there exists a constant A, independent of f, such that

$$|f(x+iy)| < Ae^{k|y|}$$

for all $x \geq 0$. From this fact and Cauchy's formula for the derivative, the result is easily seen.

LEMMA 14. *If $f(z)$ is of exponential type k in $R(z) \geq 0$, and if*

$$f(x) = O(x^r) \text{ as } x \to \infty, \qquad (-\infty < r < \infty),$$

then

$$f'(x) = O(x^r) \text{ as } x \to \infty.$$

Proof. Consider any branch of the function

$$g(z) = (z+1)^{-r} f(z+1).$$

The function $g(z)$ is regular in $R(z) \geq 0$ and is of exponential type k. Furthermore, $g(z)$ is bounded for $x \geq 0$; hence by the previous lemma there exists a constant B such that

$$|g'(x)| = |(x+1)^{-r} f'(x+1) - r(x+1)^{-r-1} f(x+1)| < B, \qquad (x \geq 0).$$

By hypothesis, $r(x+1)^{-r-1} f(x+1)$ is bounded in $x \geq 0$, hence so is $(x+1)^{-r} f'(x+1)$. This completes the proof.

THEOREM 10. *Let $f(z)$ be of exponential type k in $R(z) \geq 0$. Then,*

$$N^p\{f'(x)\} = B(k,p) N^p\{f(x)\},$$

where

$$B(k,p) = k^p \qquad \textit{if } p > 1,$$

$$B(k,p) = \frac{(p+2)2^{p+2}}{\pi k p \delta^{p-1}}(e^{k\delta p} - 1), \qquad \textit{if } p \leq 1,$$

δ *being an arbitrary positive number.*

Proof. Assume that $k < \pi$, otherwise consider $f(z/k)$. As in the previous theorem, Lemma 9 may be applied so that

$$f'(z) = f'_1(z) + f'_2(z)$$

where, by Lemmas 9 and 14, $f_1(x+iy) = O(1/x)$ as $x \to -\infty$, and $f'_2(x+iy) = O(1/x)$ as $x \to +\infty$.

As in the previous theorem, applying Lemmas 9, 10, 12 and Theorems 3, 4,

$$N^p\{f'(x)\} = N^p\{f'_1(x)\} = 2M^p\{f'_1(x)\}$$
$$\leq B(k,p)M^p\{f_1(x)\} = B(k,p)N^p\{f(x)\}.$$

Corollary. *If $f(z)$ is of minimal type in $R(z) \geq 0$, then*

$$N^p\{f(x+iy)\} = N^p\{f(x)\}, \qquad (p > 0),$$

and

$$N^p\{f'(x)\} = 0, \qquad (p > 0).$$

Proof. The first relation follows by letting k approach zero in Theorem 9.

The second relation, for $p > 1$, is proved similarly, using the relation for $p > 1$ in the preceding theorem.

In the case $p < 1$ it should be noted that in the preceding theorem,

$$\lim_{k \to 0} B(k,p) = \frac{(p+2)2^{p+2}}{\pi \delta^p}.$$

Since δ is arbitrary, by taking δ large this can be made less than any given positive quantity. This then proves the second relation for $p \leq 1$.

10. The one-sided mean of $f(z)$ at the integers.

Theorem 11. *If $f(z)$ is of exponential type k in $R(z) \geq 0$, then*

$$N_i^p\{f(x)\} \leq \frac{8(e^{\frac{1}{2}kp} - 1)}{\pi k p} N^p\{f(x)\}.$$

Proof. The proof follows that of Theorem 7 with slight modifications. The expression for $|f(j)|^p$ in Theorem 7 is valid only for $j > 0$ in this theorem, and Theorem 2 must be replaced by Theorem 9.

LEMMA 14. *Let n be an integer and let $\phi(x)$ be bounded at the negative integers. If*

$$\phi(n) \to 0 \qquad \text{as } n \to -\infty,$$

then

$$2M_i^p\{\phi(x)\} = N_i^p\{\phi(x)\}.$$

Proof. The proof is the same as for Lemma 12 with integrals replaced by sums.

LEMMA 15. *If $f(z)$ is of exponential type $k < \pi$ in $R(z) \geq 0$, and $N_i^p\{f(x)\}$ is bounded, then*

$$f(x) = O(x^{1/p}) \qquad \text{as } x \to \infty.$$

Proof. As in Lemma 7,

$$f(n) = O(n^{1/p}) \qquad \text{as } n \to \infty.$$

The function

$$g(z) = (z+1)^{-1/p} f(z+1), \qquad \text{(any branch)},$$

is regular in $R(z) \geq 0$, and bounded at the positive integers. It is of type $k < \pi$, hence $g(x)$ is bounded [13; p. 142]. The result then follows.

THEOREM 12. *If $f(z)$ is of exponential type $k < \pi$ in $R(z) \geq 0$, then*

$$N^p\{f(x)\} \leq BN_i^p\{f(x)\},$$

where B is a constant which depends on k and p only.

Proof. Assume that $N_i^p\{f(x)\}$ is bounded, otherwise there is nothing to prove.

It follows from the preceding lemmas that

$$\overline{\lim_{r\to\infty}}\, r^{-1} \log |f(r)| \leq 0.$$

Applying Lemma 9,

$$f(z) = f_1(z) + f_2(z).$$

As in Theorem 9,

$$N^p\{f(x)\} = 2M^p\{f_1(x)\}.$$

Applying Theorem 8 to $f_1(x)$, and using Lemma 14,

$$N^p\{f(x)\} \leq B2M_i{}^p\{f(x)\} = B \cdot N_i{}^p\{f(x)\}.$$

UNIVERSITY OF NEW HAMPSHIRE,
DURHAM, N. H.

BIBLIOGRAPHY

1. Bellman, R.: "A note on mean values," *Revista Universidad Nacional de Tucuman*, vol. 14, pp. 255-258.
2. Bernstein, S.: *Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle*, Paris, 1926.
3. Bernstein, V.: *Leçons sur les progrès récents de la théorie des séries de Dirichlet*, Paris, 1932.
4. Besicovitch, A. S.: *Almost periodic functions*, Cambridge, England, 1932.
5. Besicovitch, A. S.: "Analysis of conditions of generalized almost periodicity," *Acta Mathematica*, vol. 58 (1932), pp. 217-230.
6. Besicovitch, A. S., and Bohr, H.: "Almost periodicity and general trigonometric series," *Acta Mathematica*, vol. 57 (1931), pp. 203-292.
7. Boas, R. P., Jr.: "Representations for entire functions of exponential type," *Annals of Mathematics* (2), vol. 39 (1938), pp. 269-286.
8. Boas, R. P., Jr.: "Functions of exponential type, I," *Duke Mathematical Journal*, vol. 11 (1944), pp. 9-15.
9. Boas, R. P., Jr.: "Entire functions bounded on a line," *Duke Mathematical Journal*, vol. 6 (1940), pp. 148-169.
10. Bohr, H.: see Besicovitch and Bohr.
11. Cartwright, M. L.: "On certain integral functions of order one," *Quarterly Journal of Mathematics* (Oxford series), vol. 7 (1936), pp. 46-55.
12. Doetsch, G.: *Theorie und Anwendung der Laplace Transformation*, vol. 1, Berlin, 1937.
13. Duffin, R. J. and Schaeffer, A. C.: "Power series with bounded coefficients," *American Journal of Mathematics*, vol. 67 (1945), pp. 141-154.
14. Hardy, G. H.: "The mean value of the modulus of an analytic function," *Proceedings of the London Mathematical Society* (2), vol. 14 (1915), pp. 269-277.
15. Hardy, G. H., Ingham, A. E. and Pólya, G.: "Theorems concerning mean values of analytic functions," *Proceedings of the Royal Society, A*, vol. 113 (1927), pp. 542-569.
16. Hardy, G. H., Littlewood, J. E and Pólya, G.: *Inequalities*, Cambridge, England, 1934.

17. Hobson, E. W.: *Theory of functions of a real variable*, vol. II, Cambridge, England, 1926.

18. Ingham, A. E.: see Hardy, Ingham and Pólya.

19. Littlewood, J. E.: see Hardy, Littlewood and Pólya.

20. Levinson, N.: *Gap and density theorems*, American Mathematical Society Colloquium Publications, New York, 1940.

21. Macintyre, A. J.: "Laplace's transformation and integral functions," *Proceedings of the London Mathematical Society* (2), vol. 45 (1938-39), pp. 1-20.

22. Nevanlinna, R.: "Untersuchungen über den Picard'schen Satz, *Acta Societatis Scientiarum Fennicae*, vol. 50 (1924), pp. 1-42.

23. Pfluger, A.: "On analytic functions bounded at the lattice points," *Proceedings of the London Mathematical Society* (2), vol. 42 (1936-37), pp. 305-315.

24. Plancherel, M. and Pólya, G.: "Fonctions entières et intégrales de Fourier multiples (part two)," *Commentarii Mathematici Helvitici*, vol. 10 (1937-38), pp. 110-163.

25. Pólya, G.: Problem 105, *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 40 (1931), p. 80.

26. Pólya, G.: "Untersuchungen über Lücken und Singularitäten von Potenzreihen," *Mathematische Zeitschrift*, vol. 29 (1929), pp. 549-640.

27. Pólya, G., and Szegö, G.: *Aufgaben und Lehrsätze aus der Analysis* I-II, Berlin (1925).

28. Pólya, G.: see Hardy, Ingham and Pólya; Hardy, Littlewood and Pólya; Plancherel and Pólya.

29. Schaeffer, A. C.: see Duffin and Schaeffer.

30. Szegö, G.: see Pólya and Szegö.

31. Titchmarsh, E. C.: *Theory of functions*, Oxford (1939).

32. Titchmarsh, E. C.: *Introduction to the theory of Fourier integrals*, Oxford (1937).

33. Tschakaloff, L.: "Zweite Losung der Aufgabe 105, *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 43 (1934), p. 11.

34. Valiron, G.: "Sur la formule d'interpolation de Lagrange," *Bulletin des Sciences Mathématiques* (2), vol. 49 (1925), pp. 181-192, 203-224.

35. Valiron, G.: *Lectures on the general theory of integral functions*, Toulouse, 1932.

QUASI-MONOTONE SERIES.*

By Otto Szász.

1. We shall call a sequence $\{a_n\}$ of positive numbers quasi-monotone if for some constant $\alpha \geqq 0$

$$a_{n+1} \leqq a_n(1+\alpha/n) \text{ for all } n > n_0(\alpha). \tag{1.1}$$

For $\alpha = 0$ and $n_0 = 0$ we have a monotone non-increasing sequence. If the terms a_n of an infinite series Σa_n form a quasi-monotone sequence, the series will be called quasi-monotone. Many known theorems on monotone series can be generalized to the larger class of quasi-monotone series. For some such results see [4], § 4; [5], § 4; [6], § 5,[1] and references given there.

In this note we give some properties of such series, which include results due to R. W. Hamming [2] and to A. Dvoretzky [1].

2. Theorem 1. *Let $\lambda_0 < \lambda_1 < \cdots$ be a strictly increasing sequence of positive integers, such that*

$$\lambda_{n+1} - \lambda_n = O(\lambda_n - \lambda_{n-1}), \qquad \textit{as } n \to \infty. \tag{2.1}$$

If $\{a_n\}$ satisfies the assumption (1.1) then the two series

$$\Sigma a_n, \textit{ and } \Sigma(\lambda_n - \lambda_{n-1})a_{\lambda_n}$$

are either both convergent or both divergent.

The special case $\alpha = 0$ is well known. From (1.1)

$$a_{n-1} \geqq (1+\alpha/(n-1))^{-1}a_n, \qquad a_{n-2} \geqq (1+\alpha/(n-2))^{-2}a_n, \cdots$$

hence, for $k < n$

$$\sum_{\nu=1}^{k} a_{n-\nu} \geqq k(1+\alpha/(n-k))^{-k}a_n,$$

or

$$ka_n \leqq (1+\alpha/(n-k))^k \sum_{\nu=1}^{k} a_{n-\nu} < (\exp k\alpha/(n-k)) \sum_{\nu=1}^{k} a_{n-\nu}. \tag{2.2}$$

* Received April 24, 1947; Presented to the American Mathematical Society, April 26, 1947.

[1] See the bibliography at the end of the paper.

It follows from (2.1) that $\lambda_{n+1} = O(\lambda_n)$, and from (2.2), replacing n by λ_n, k by $\lambda_n - \lambda_{n-1}$,

$$(2.3) \qquad (\lambda_n - \lambda_{n-1}) a_{\lambda_n} < (\exp \alpha(\lambda_n - \lambda_{n-1})/\lambda_{n-1}) \sum_{\lambda_{n-1}}^{\lambda_n - 1} a_\nu = O \sum_{\lambda_{n-1}}^{\lambda_n - 1} a_\nu.$$

Again, from (1.1)

$$a_{n+\nu} \leq (1 + \alpha/n)^\nu a_n, \qquad \nu = 0, 1, 2, \cdots,$$

hence

$$\sum_{\nu=1}^{k} a_{n+\nu} \leq k a_n (1 + \alpha/n)^k < k a_n \exp k\alpha/n,$$

and

$$(2.4) \qquad (\lambda_{n+1} - \lambda_n) a_{\lambda_n} > \exp\left(-\alpha(\lambda_{n+1} - \lambda_n)/\lambda_n\right) \sum_{1+\lambda_n}^{\lambda_{n+1}} a_\nu > c \sum_{1+\lambda_n}^{\lambda_{n+1}} a_\nu,$$

c a constant; (2.3) and (2.4) yield Theorem 1.

COROLLARY. *If Σb_n is another quasi-monotone series and if*

$$(2.5) \qquad a_{\lambda_n} \geq c b_{\lambda_n}, \qquad n = 1, 2, \cdots,$$

then $\Sigma a_n < \infty$ implies $\Sigma b_n < \infty$.

Thus if Σb_n is divergent, then (2.5) cannot hold; for $\lambda_n = k^n$ this yields Theorem 1 of Dvoretzky.

Note that if $\Sigma a_n < \infty$ then, (2.2) with $k = [n/2]$ yields $n a_n \to 0$.

3. Analogous results hold for infinite integrals.

THEOREM 2. *Assume that*

$$(3.1) \qquad 0 < a(x + y) \leq (1 + \alpha/x) a(x) \ \textit{for}\ x \geq 1, \qquad 0 < y < 1,$$

and that

$$a(n) = a_n, \qquad n = 1, 2, 3, \cdots,$$

then the series Σa_n and the integral $\int_1^\infty a(x)\,dx$ either both are convergent or both divergent.

From (3.1) for $\nu \geq 1$

$$a(\nu + t) \leq (1 + \alpha/\nu) a_\nu, \qquad 0 < t < 1,$$

$$(3.2) \qquad \int_\nu^{\nu+1} a(x)\,dx = \int_0^1 a(\nu + t)\,dt \leq a_\nu(1 + \alpha/\nu) \leq (1 + \alpha) a_\nu;$$

furthermore

$$a_{\nu+1} \leq a(\nu+t)(1+\alpha/(\nu+t)) < a(\nu+t)(1+\alpha),$$

hence

$$(3.3) \qquad a_{\nu+1} \leq (1+\alpha) \int_0^1 a(\nu+t)\,dt.$$

Thus

$$1/(1+\alpha) \int_1^{n+1} a(x)\,dx \leq \sum_1^n a_\nu \leq a_1 + (1+\alpha) \int_1^n a(x)\,dx.$$

This proves Theorem 2.

Note that if $\Sigma a_n < \infty$, then $na_n \to 0$; furthermore

$$\nu a(\nu+t) \leq (\nu+\alpha) a_\nu;$$

thus if $\int_1^\infty a(x)\,dx < \infty$, then $xa(x) \to 0$.

4. Assume again (3.1) and suppose that $\int_0^\infty a(x)\,dx < \infty$. Let $0 < h < 1$, then

$$\int_{\nu h}^{(\nu+1)h} a(x)\,dx = \int_0^h a(\nu h + y)\,dy \leq \int_0^h a(\nu h)(1+\alpha/\nu h)\,dy = a(\nu h)(h+\alpha/\nu),$$

$$\int_{nh}^{(m+1)h} a(x)\,dx \leq h \sum_n^m a(\nu h) + \alpha \sum_n^m a(\nu h)/\nu, \qquad n = 1, 2, 3, \cdots.$$

Similarly, if $\nu h \geq 1$, from

$$a(\nu h + h) \leq a(\nu h + y)(1+\alpha/(\nu h + y)) \leq a(\nu h + y)(1+\alpha/\nu h), \qquad 0 < y < h,$$

$$h \sum_n^m a((\nu+1)h) \leq \int_{nh}^{(m+1)h} a(x)\,dx + \alpha/h \sum_n^m 1/\nu \int_{\nu h}^{(\nu+1)h} a(x)\,dx, \qquad nh \geq 1.$$

Let $\lambda > 1$, and $n = 1 + [h^{-1}]$, $m = [\lambda/h]$, then, using $xa(x) = o(1)$,

$$h \sum_n^m a(\nu h) \to \int_1^\lambda a(x)\,dx, \qquad h \to 0,$$

and

$$\sum_m^\infty a(\nu h)\nu^{-1} = o(1/h \sum_m^\infty 1/\nu^2) = o(1/hm) = o(1/\lambda) = o(1), \qquad h \to 0.$$

Similarly

$$1/h \sum_m^\infty 1/\nu \int_{\nu h}^{(\nu+1)h} a(x)\,dx = (1/h)\, o \sum_m^\infty 1/\nu^2 = o(1/\lambda) = o(1).$$

Given $\epsilon > o$, choose $\lambda = 1 + \epsilon$, and $h < h(\epsilon)$ so small that $o(1) < \epsilon$; it then follows that

$$\left| \int_{nh}^{\infty} a(x)\,dx - h \sum_{\nu h > 1} a(\nu h) \right| < \epsilon, \qquad n = 1 + [h^{-1}],$$

hence

$$\lim_{h \to 0} h \sum_{\nu h > 1} a(\nu h) = \int_{1}^{\infty} a(x)\,dx.$$

Furthermore, clearly

$$h \sum_{\nu h \leq 1} a(\nu h) \to \int_{0}^{1} a(x)\,dx,$$

hence, finally

$$\lim_{h \to 0} h \sum_{\nu=1}^{\infty} a(\nu h) = \int_{0}^{\infty} a(x)\,dx.$$

This proves

THEOREM 3. *If* (3.1) *holds, and if* $a(x)$ *is Riemann integrable then*

$$h \sum_{\nu=1}^{\infty} a(\nu h) \to \int_{0}^{\infty} a(x)\,dx, \qquad \text{as } h \to 0,$$

whenever the integral exists.

For monotone $a(x)$ see [3], vol. 1, p. 41.

THE UNIVERSITY OF CINCINNATI.

REFERENCES

1. Aryeh Dvoretzky, "On monotone series," *American Journal of Mathematics*, vol. 70 (1948), pp. 167-173.
2. R. W. Hamming, "Monotone series," *American Journal of Mathematics*, vol. 68 (1946), pp. 133-136.
3. G. Pólya and G. Szegö, *Aufgaben und Lehrsatze aus der Analysis*, Berlin, 1925.
4. Otto Szász, "On uniform convergence of Fourier series," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 587-595.
5. Otto Szász, "On uniform convergence of trigonometric series," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 856-867.
6. Otto Szász, "On the absolute convergence of trigonometric series," *Annals of Mathematics*, vol. 47 (1946), pp. 213-220.

ON THE PARTIAL PRODUCTS OF INFINITE PRODUCTS OF ALEPHS.*

By FREDERICK BAGEMIHL.

Throughout this paper we shall assume the following:[1]

1. λ is a transfinite limit-number.
2. For every $\xi < \lambda$, $\mathfrak{a}_\xi$ is a transfinite cardinal number.
3. $\mathfrak{s} = \sum_{\xi<\lambda} \mathfrak{a}_\xi$, $\mathfrak{p} = \prod_{\xi<\lambda} \mathfrak{a}_\xi$.
4. $\mathfrak{p}_0 = 1$; $\mathfrak{p}_\xi = \prod_{\iota<\xi} \mathfrak{a}_\iota$ for every ξ such that $0 < \xi < \lambda$. We call the $\mathfrak{p}_\xi$ the "partial products" of $\prod_{\xi<\lambda} \mathfrak{a}_\xi$.
5. For any α, $\bar{\alpha}$ is the cardinal number of the set of all ordinal numbers less than α.

A generalization by Jourdain [2] of a theorem of König [3] can be expressed as follows:

THEOREM 1. (*König-Jourdain.*)

$$\mathfrak{s}^{\bar{\lambda}} = \mathfrak{p}^{\bar{\lambda}}.$$

Tarski has proved [6, pp. 6-7] that if $\{\sigma_\xi\}_{\xi<\omega\delta}$ is an increasing sequence of ordinal numbers, and if $\lim_{\xi<\omega\delta} \sigma_\xi = \lambda$, then

$$\prod_{\xi<\omega\delta} \aleph_{\sigma_\xi} = \aleph_\lambda^{\overline{\lambda\omega\delta}}.$$

If we make use of Theorem 1 and the fact that

$$\aleph_\lambda = \sum_{\xi<\omega\delta} \aleph_{\sigma_\xi}$$

(see, *e.g.*, [4, p. 135]), we can restate Tarski's theorem with the aid of our notation thus:

* Received April 29, 1947.

[1] Small German and Greek letters, with or without subscripts, denote cardinal and ordinal numbers, respectively. For an explanation of our notation and terminology the reader may consult [1], [5], or [6]. Numbers in brackets refer to the bibliography at the end of the paper.

THEOREM 2. (*Tarski.*) *If*

(1) $$\mathfrak{a}_\alpha < \mathfrak{a}_\beta \qquad \textit{whenever } \alpha < \beta < \lambda,$$

and if

(2) $$\lambda = \omega^\delta,$$

then

(3) $$\mathfrak{p} = \mathfrak{p}^{\bar{\lambda}}.$$

Now, Tarski has asked [6, p. 13] (in an equivalent form): Is it true or false that Theorem 2 remains valid when the hypothesis (2) is removed? This is still an open question, and has been decided, in the affirmative, only for certain special cases.[2] We shall first prove

THEOREM 3. *If*

(4) $$\mathfrak{a}_\alpha \leqq \mathfrak{a}_\beta \qquad \textit{whenever } \alpha < \beta < \lambda,$$

and (2), *then* (3).

Proof. We give a proof which is entirely analogous to Tarski's proof of Theorem 2, and we refer the reader to the latter author's paper for the notions and facts therein which we shall employ but which we shall not restate here.

Let $\tau_{\xi,\eta} = \xi$ for $\xi < \lambda$ and $\eta < \lambda$. Then

(5) $$\mathfrak{p}^{\bar{\lambda}} = \prod_{\xi<\lambda} \mathfrak{a}_\xi^{\bar{\lambda}} = \prod_{\zeta<\lambda} \Big(\prod_{\eta<\lambda} \mathfrak{a}_{\tau_{\xi,\eta}} \Big) = \prod_{\zeta<\lambda} \Big(\prod_{\xi\#\eta=\zeta} \mathfrak{a}_{\tau_{\xi,\eta}} \Big).$$

Every product

$$\prod_{\xi\#\eta=\zeta} \mathfrak{a}_{\tau_{\xi,\eta}}, \qquad \zeta < \lambda,$$

has a finite number of factors, and because of (4) the factor $\mathfrak{a}_{\tau_{\zeta,0}} = \mathfrak{a}_\zeta$ is not less than any other factor appearing in this product. Hence,

$$\prod_{\xi\#\eta=\zeta} \mathfrak{a}_{\tau_{\xi,\eta}} = \mathfrak{a}_\zeta \qquad \text{for } \zeta < \lambda;$$

and substituting this in (5) we obtain $\mathfrak{p}^{\bar{\lambda}} = \mathfrak{p}$.

Thus we see that Theorem 2 is valid if (1) is replaced by (4). This suggests a question analogous to Tarski's, namely: Does Theorem 3 hold if

---

[2] See [6, pp. 11, 13-14]. Some results which are related to this problem are stated in [7].

hypothesis (2) is removed? The following example shows that this cannot be answered in the affirmative:[3]

Let $\lambda = \omega_1 + \omega$, $\mathfrak{a}_\xi = \aleph_0$ for $\xi < \omega_1$, and[4] $\mathfrak{a}_{\omega_1+\xi} = \aleph_{\pi(\omega_1)}$ for $\xi < \omega$. Then

$$\mathfrak{p} = (\prod_{\xi<\omega_1} \aleph_0)(\prod_{\xi<\omega} \aleph_{\pi(\omega_1)}) = 2^{\aleph_1} \cdot \aleph_{\pi(\omega_1)}^{\aleph_0} = \aleph_{\pi(\omega_1)},$$

where the last equality results from the relation $\aleph_{\pi(\omega_1)} > 2^{\aleph_1}$, [6, p. 9, Theorem II, 1)], and the fact that $cf(\omega_1) = 1$. On the other hand, noting that $\bar{\lambda} = \aleph_1$, we obtain

$$\mathfrak{p}^{\bar{\lambda}} = \aleph_{\pi(\omega_1)}^{\aleph_1} = 2^{\aleph_{\pi(\omega_1)}},$$

where the last equality is a consequence of [6, p. 9, Theorem II, 2)]. Hence, $\mathfrak{p}^{\bar{\lambda}} > \mathfrak{p}$.

What, then, can be expected in the way of a conclusion resembling (3), if only (4) is assumed? We prove

THEOREM 4. *If* (4), *then*

$$\mathfrak{p} = \prod_{\xi<\lambda} \mathfrak{p}_\xi. \tag{6}$$

Proof. Let the normal form of λ be: $\lambda = \kappa_1 + \kappa_2 + \cdots + \kappa_n$, where n is a natural number, $\kappa_i = \omega^{\delta_i}$ for $i = 1, 2, \cdots, n$, and

$$\delta_1 \geqq \delta_2 \geqq \cdots \geqq \delta_n > 0. \tag{7}$$

Then

$$\mathfrak{p} = (\prod_{\xi<\kappa_1} \mathfrak{a}_\xi)\ (\prod_{\xi<\kappa_2} \mathfrak{a}_{\kappa_1+\xi}) \cdots (\prod_{\xi<\kappa_n} \mathfrak{a}_{\kappa_1+\kappa_2+\cdots+\kappa_{n-1}+\xi}), \tag{8}$$

(where, if $n = 1$, only the product in the first set of parentheses is present). According to Theorem 3, the right-hand side of (8) is equal to

$$(\prod_{\xi<\kappa_1} \mathfrak{a}_\xi)^{\bar{\kappa}_1} (\prod_{\xi<\kappa_2} \mathfrak{a}_{\kappa_1+\xi})^{\bar{\kappa}_2} \cdots (\prod_{\xi<\kappa_n} \mathfrak{a}_{\kappa_1+\kappa_2+\cdots+\kappa_{n-1}+\xi})^{\bar{\kappa}_n},$$

which in turn equals

$$(\prod_{\xi<\kappa_1} \mathfrak{a}_\xi)^{\overline{\kappa_1+\cdots+\kappa_n}} (\prod_{\xi<\kappa_2} \mathfrak{a}_{\kappa_1+\xi})^{\overline{\kappa_2+\cdots+\kappa_n}} \cdots (\prod_{\xi<\kappa_n} \mathfrak{a}_{\kappa_1+\kappa_2+\cdots+\kappa_{n-1}+\xi})^{\bar{\kappa}_n}, \tag{9}$$

because of (7). But since $\overline{\rho + \sigma} = \bar{\rho} + \bar{\sigma}$ for every ρ and σ, (9) is the same as

[3] The author wishes to express his thanks to the referee for suggesting a simplification of the original version of this example.

[4] For a definition of $\pi(\gamma)$ for any ordinal number γ, see [6, p. 9].

$$(10)\quad (\prod_{\xi<\kappa_1} \mathfrak{a}_\xi)^{\bar{\kappa}_1}(\prod_{\xi<\kappa_1+\kappa_2} \mathfrak{a}_\xi)^{\bar{\kappa}_2}\cdots(\prod_{\xi<\kappa_1+\kappa_2+\cdots+\kappa_n} \mathfrak{a}_\xi)^{\bar{\kappa}_n} = \mathfrak{p}_{\kappa_1}^{\bar{\kappa}_1}\cdot\mathfrak{p}_{\kappa_1+\kappa_2}^{\bar{\kappa}_2}\cdots\mathfrak{p}_{\kappa_1+\kappa_2+\cdots+\kappa_n}^{\bar{\kappa}_n}.$$

From the definition of $\mathfrak{p}_\xi$ it is evident that $\mathfrak{p}_\alpha \leqq \mathfrak{p}_\beta$ whenever $\alpha < \beta < \lambda$, so that

$$\mathfrak{p}_{\kappa_1}^{\bar{\kappa}_1}\cdot\mathfrak{p}_{\kappa_1+\kappa_2}^{\bar{\kappa}_2}\cdots\mathfrak{p}_{\kappa_1+\kappa_2+\cdots+\kappa_n}^{\bar{\kappa}_n} \geqq (\prod_{\xi<\kappa_1}\mathfrak{p}_\xi)(\prod_{\xi<\kappa_2}\mathfrak{p}_{\kappa_1+\xi})\cdots(\prod_{\xi<\kappa_n}\mathfrak{p}_{\kappa_1+\kappa_2+\cdots+\kappa_{n-1}+\xi})$$
$$= \prod_{\xi<\lambda}\mathfrak{p}_\xi.$$

Thus $\mathfrak{p} \geqq \prod_{\xi<\lambda}\mathfrak{p}_\xi$; and since obviously $\mathfrak{p} \leqq \prod_{\xi<\lambda}\mathfrak{p}_\xi$, we must have (6).

Theorem 4 is of course true, *a fortiori,* for every transfinite limit-number λ, if hypothesis (4) is replaced by (1).

We give an example to show that Theorem 4 is not valid if the hypothesis (4) is dropped:

Let $\lambda = \omega$. Set $\mathfrak{a}_0 = \aleph_{\pi(\omega)}$; $\mathfrak{a}_\xi = \aleph_0$ for $0 < \xi < \lambda$. From the definition[4] of $\aleph_{\pi(\omega)}$ it is obvious that $\aleph_{\pi(\omega)} > 2^{\aleph_0}$. Now

$$\mathfrak{p} = \aleph_{\pi(\omega)}\cdot\prod_{\xi<\omega}\aleph_0 = \aleph_{\pi(\omega)}\cdot 2^{\aleph_0} = \aleph_{\pi(\omega)}.$$

On the other hand, $\mathfrak{p}_0 = 1$, and $\mathfrak{p}_\xi = \aleph_{\pi(\omega)}$ for $0 < \xi < \lambda$, so that according to [6, p. 9, Theorem II, 2)]

$$\prod_{\xi<\lambda}\mathfrak{p}_\xi = \aleph_{\pi(\omega)}^{\aleph_0} = 2^{\aleph_{\pi(\omega)}} > \mathfrak{p}.$$

Tarski has stated [7, Theorem 97] a result which we can express in the following form:

If (1), *then there exists a non-zero remainder,* ρ, *of* λ, *such that* $\mathfrak{p} = \mathfrak{p}^{\bar{\rho}}$.

We show that the hypothesis (1) can be replaced by (4):

THEOREM 5. *If* (4), *and if the normal form of* λ *is:* $\lambda = \kappa_1 + \kappa_2 + \cdots + \kappa_n$, *where n is a natural number,* $\kappa_i = \omega^{\delta_i}$ *for* $i = 1, 2, \cdots, n$, *and* $\delta_1 \geqq \delta_2 \geqq \cdots \geqq \delta_n > 0$, *then*

$$(11)\qquad \mathfrak{p} = \mathfrak{p}^{\bar{\kappa}_n}.$$

Proof. If $n = 1$, Theorem 5 is simply Theorem 3.

Suppose $n > 1$. Then (11) follows immediately from

$$\mathfrak{p} = \mathfrak{p}_{\kappa_1}^{\bar{\kappa}_1} \cdot \mathfrak{p}_{\kappa_1+\kappa_2}^{\bar{\kappa}_2} \cdots \mathfrak{p}_{\kappa_1+\kappa_2+\cdots+\kappa_n}^{\bar{\kappa}_n} \geqq \mathfrak{p}_{\kappa_1+\kappa_2+\cdots+\kappa_n}^{\bar{\kappa}_n} = \mathfrak{p}^{\bar{\kappa}_n} \geqq \mathfrak{p},$$

where the first equality is obtained from (10) and the argument preceding it in the proof of Theorem 4.

THE UNIVERSITY OF ROCHESTER.

BIBLIOGRAPHY

1. F. Hausdorff, *Mengenlehre*, 3d ed. (reprint), New York, 1944.

2. Ph. E. B. Jourdain, "On infinite sums and products of cardinal numbers," *Quarterly Journal of Pure and Applied Mathematics*, vol. 39 (1908), p. 380, footnote †.

3. J. König, "Zum Kontinuum-Problem," *Mathematische Annalen*, vol. 60 (1905), p. 178.

4. A. Schoenflies, *Entwickelung der Mengenlehre und ihrer Anwendungen*, vol. I, Leipzig and Berlin, 1913.

5. W. Sierpiński, *Leçons sur les nombres transfinis*, Paris, 1928.

6. A. Tarski, "Quelques théorèmes sur les alephs," *Fundamenta Mathematicae*, vol. 7 (1925), pp. 1-14.

7. ———, (and A. Lindenbaum), "Communication sur les recherches de la théorie des ensembles," *Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie*, vol. 19 (1926), pp. 314-315.

A BILINEAR INTEGRAL IDENTITY FOR HARMONIC FUNCTIONS.*

By WILLIAM GUSTIN.

In this note we show that any two functions harmonic in open subsets of a euclidean space satisfy a certain bilinear integral identity. We use this identity and the associated quadratic integral identity to give new proofs of several fundamental theorems of harmonic function theory. In particular we show, without the use of series expansions, that a function harmonic in a connected open set D and vanishing over some non-null open subset of D must vanish throughout D.

Let E be a euclidean space of dimension $\nu \geqq 2$. The elements of E will be called points or vectors depending upon context; all points or point sets hereafter mentioned will lie in E. We use small Roman italics to denote points or vectors in E, capital Roman italics to denote point sets in E, and small Greek italics to denote real numbers or real valued functions.

Let ϕ_1 be a function harmonic in an open set D_1 of E and let ϕ_2 be a function harmonic in an open set D_2 of E. For every two non-negative real numbers ρ_1 and ρ_2 such that the closed ρ_1 sphere about the point q_1 lies in D_1 and the closed ρ_2 sphere about the point q_2 lies in D_2 we consider a bilinear integral of ϕ_1 and ϕ_2 defined as follows. Any unit vector x determines a direction or ray from each of the points q_1 and q_2. The ray from q_1 pierces the surface of the closed ρ_1 sphere about q_1 in the point $q_1 + \rho_1 x$; and the ray from q_2 pierces the surface of the closed ρ_2 sphere about q_2 in the point $q_2 + \rho_2 x$. Thus the product

$$\phi_1(q_1 + \rho_1 x)\phi_2(q_2 + \rho_2 x)$$

is defined for every unit vector x. This product may be integrated with respect to surface measure over the set S of all unit vectors in E, that is, over the surface of the unit sphere about the origin. The resulting bilinear integral will be denoted by

$$\int_{S_x} \phi_1(q_1 + \rho_1 x)\phi_2(q_2 + \rho_2 x),$$

* Received March 6, 1947.

the subscript x specifying that the unit vector x is the variable point of integration. We now investigate the dependence of this integral upon the radii ρ_1 and ρ_2.

THEOREM 1. *If the function ϕ_1 is harmonic in an open set D_1 of E and the function ϕ_2 is harmonic in an open set D_2 of E, then for each point q_1 of D_1 and each point q_2 of D_2 the bilinear integral identity*

$$(1) \qquad \int_{S_a} \phi_1(q_1 + \alpha_1 a)\phi_2(q_2 + \alpha_2 a) = \int_{S_b} \phi_1(q_1 + \beta_1 b)\phi_2(q_2 + \beta_2 b)$$

holds for all non-negative real numbers α_1, α_2, β_1, β_2 such that $\alpha_1\alpha_2 = \beta_1\beta_2$ and such that the closed α_1 and β_1 spheres about q_1 lie in D_1 and the closed α_2 and β_2 spheres about q_2 lie in D_2.

Proof. Let $\gamma_1 > 0$ be the radius of the largest open sphere in D_1 about q_1 and let $\gamma_2 > 0$ be the radius of the largest open sphere in D_2 about q_2. Clearly insofar as the proof of the theorem is concerned the points q_1 and q_2 may be taken as the origin. We then define

$$\zeta(\rho_1, \rho_2) = \int_{S_x} \phi_1(\rho_1 x)\phi_2(\rho_2 x)$$

for all real numbers ρ_1 and ρ_2 such that $0 \leqq \rho_1 < \gamma_1$ and $0 \leqq \rho_2 < \gamma_2$. We are to show that if $\zeta(\alpha_1, \alpha_2)$ and $\zeta(\beta_1, \beta_2)$ are defined and if $\alpha_1\alpha_2 = \beta_1\beta_2$ then

$$(2) \qquad \zeta(\alpha_1, \alpha_2) = \zeta(\beta_1, \beta_2).$$

By interchanging, if necessary, the symbols α and β we may suppose that $0 \leqq \alpha_1 \leqq \beta_1$. Moreover, since ζ is a continuous function of its arguments, we need prove (2) only for α's and β's such that $0 < \alpha_1 < \beta_1$. We shall give two proofs. The first is based on the Poisson integral formula and the second is based on Green's bilinear integral identity.

Let ϕ be a function harmonic in the open sphere in E of radius $\beta > 0$ about the origin and continuous in the closure of this sphere. Any point in the open β sphere about the origin is of the form αa where a is a unit vector and $0 \leqq \alpha < \beta$. The value of the harmonic function ϕ at such a point αa is determined by the values of ϕ on the surrounding surface of the β sphere about the origin according to the following integral expression, known as the Poisson integral formula,

$$\phi(\alpha a) = \int_{S_b} \psi(\alpha a, \beta b)\phi(\beta b).$$

The function ψ in the integrand is called the Poisson kernel and is defined for all real numbers α and β such that $0 \leqq \alpha < \beta$ and for all unit vectors a and b as

$$\psi(\alpha a, \beta b) = \frac{\beta^{\nu-2}(\beta^2 - \alpha^2)}{\sigma[\alpha^2 + \beta^2 - 2\alpha\beta \cos\theta(a, b)]^{\nu/2}},$$

where σ is the surface measure of the unit spherical surface S and $\theta(a, b)$ is the angle between the two unit vectors a and b.

Now $\alpha_1\alpha_2 = \beta_1\beta_2$, so we may define the ratio $\kappa = \alpha_1/\beta_2 = \beta_1/\alpha_2$. Consequently $\alpha_1 = \kappa\beta_2$ and $\beta_1 = \kappa\alpha_2$, and the inequality $0 < \alpha_1 < \beta_1$ implies the inequality $0 < \beta_2 < \alpha_2$. The Poisson kernels $\psi(\alpha_1 a, \beta_1 b)$ and $\psi(\beta_2 b, \alpha_2 a)$ are then defined. Moreover, since ψ is a homogeneous function of degree zero and since $\theta(a, b) = \theta(b, a)$, we have

$$\psi(\alpha_1 a, \beta_1 b) = \psi(\kappa\beta_2 a, \kappa\alpha_2 b) = \psi(\beta_2 a, \alpha_2 b) = \psi(\beta_2 b, \alpha_2 a).$$

The integral $\zeta(\alpha_1, \alpha_2)$ may be developed according to the Poisson formula for the value of ϕ_1 at the point $\alpha_1 a$ on the surface of the α_1 sphere about the origin as determined by the values of ϕ_1 on the surrounding surface of the β_1 sphere about the origin; thus

$$\zeta(\alpha_1, \alpha_2) = \int_{S_a} \int_{S_b} \psi(\alpha_1 a, \beta_1 b)\phi_1(\beta_1 b) \cdot \phi_2(\alpha_2 a).$$

Substituting the kernel $\psi(\beta_2 b, \alpha_2 a)$ for the kernel $\psi(\alpha_1 a, \beta_1 b)$ in this iterated integral and interchanging the order of iteration, we obtain

$$\zeta(\alpha_1, \alpha_2) = \int_{S_b} \int_{S_a} \psi(\beta_2 b, \alpha_2 a)\phi_2(\alpha_2 a) \cdot \phi_1(\beta_1 b).$$

Now the inner integral with respect to a is the Poisson formula for the value of ϕ_2 at the point $\beta_2 b$ on the surface of the β_2 sphere about the origin as determined by the values of ϕ_2 on the surrounding surface of the α_2 sphere about the origin. Consequently the above iterated integral has the value

$$\int_{S_b} \phi_1(\beta_1 b)\phi_2(\beta_2 b) = \zeta(\beta_1, \beta_2).$$

This completes the first proof of the theorem.

The essential content of the theorem is that the integral $\zeta(\rho_1, \rho_2)$ depends

only upon the value of the product $\rho_1\rho_2$ and not upon the particular values of ρ_1 and ρ_2. We base our second proof of the theorem upon this remark.

Let ρ be any real number in the open interval $0 < \rho < \sqrt{\gamma_1\gamma_2}$; and for such a ρ let λ be any real number in the open interval $\rho/\gamma_2 < \lambda < \gamma_1/\rho$. Then $0 < \lambda\rho < \gamma_1$ and $0 < \lambda^{-1}\rho < \gamma_2$; and we define the function $\eta(\lambda, \rho)$ as follows:

$$\eta(\lambda, \rho) = \zeta(\lambda\rho, \lambda^{-1}\rho) = \int_{S_x} \phi_1(\lambda\rho x)\phi_2(\lambda^{-1}\rho x). \tag{3}$$

We shall show that this function is independent of λ. Consequently, since $\alpha_1\alpha_2 = \beta_1\beta_2$, we shall have proved that

$$\zeta(\alpha_1, \alpha_2) = \eta(\sqrt{\alpha_1/\alpha_2}, \sqrt{\alpha_1\alpha_2}) = \eta(\sqrt{\beta_1/\beta_2}, \sqrt{\beta_1\beta_2}) = \zeta(\beta_1, \beta_2).$$

Differentiating (3) under the integral we obtain

$$(\lambda\delta/\delta\lambda)\eta(\lambda, \rho) = \int_{S_x} [(\delta\lambda/\delta\lambda)\phi_1(\lambda\rho x) \cdot \phi_2(\lambda^{-1}\rho x) + \phi_1(\lambda\rho x) \cdot (\lambda\delta/\delta\lambda)\phi_2(\lambda^{-1}\rho x)].$$

The operators $\lambda\delta/\delta\lambda$ and $-\lambda^{-1}\delta/\delta\lambda^{-1}$ are equivalent when applied to a differentiable function of λ; the operators $\lambda\delta/\delta\lambda$ and $\rho\delta/\delta\rho$ are equivalent when applied to a differentiable function of $\lambda\rho$; and the operators $\lambda^{-1}\delta/\delta\lambda^{-1}$ and $\rho\delta/\delta\rho$ are equivalent when applied to a differentiable function of $\lambda^{-1}\rho$. Therefore

$$(\lambda\delta/\delta\lambda)\eta(\lambda, \rho) = \rho \int_{S_x} [(\delta/\delta\rho)\phi_1(\lambda\rho x) \cdot \phi_2(\lambda^{-1}\rho x) - \phi_1(\lambda\rho x) \cdot (\delta/\delta\rho)\phi_2(\lambda^{-1}\rho x)].$$

An expansion of modulus λ^{-1} about the origin is a conformal mapping which transforms the harmonic function ϕ_1 defined in the open γ_1 sphere about the origin into another harmonic function ϕ'_1 defined in the open $\lambda^{-1}\gamma_1 > \rho$ sphere about the origin; and an expansion of modulus λ about the origin is a conformal mapping which transforms the harmonic function ϕ_2 defined in the open γ_2 sphere about the origin into another harmonic function ϕ'_2 defined in the open $\lambda\gamma_2 > \rho$ sphere about the origin. Thus we have

$$(\lambda\delta/\delta\lambda)\eta(\lambda, \rho) = \rho \int_{S_x} [(\delta/\delta\rho)\phi'_1(\rho x) \cdot \phi'_2(\rho x) - \phi'_1(\rho x) \cdot (\delta/\delta\rho)\phi'_2(\rho x)] = 0.$$

The integral vanishes in consequence of Green's bilinear integral identity, for the functions ϕ'_1 and ϕ'_2 are harmonic in open sets containing the closed ρ sphere about the origin and the directional differentiation $\delta/\delta\rho$ is taken along the outward normal to the surface of this sphere. The function $\eta(\lambda, \rho)$,

being defined for λ in the open interval $\rho/\gamma_2 < \lambda < \gamma_1/\rho$, is then independent of λ. This concludes the second proof of the theorem.

We first apply Theorem 1 to a pair of harmonic functions, one of which is homogeneous.

THEOREM 2. *If ϕ is harmonic in an open set D of E and ω is harmonic in E and homogeneous about the origin of degree μ, then the integral expression*

$$\rho^{-\mu} \int_{S_x} \phi(q + \rho x)\omega(x),$$

defined for every point q of D and for every positive real number ρ such that the closed ρ sphere about q lies in D, is independent of ρ and is harmonic in D as a function of q.

Proof. Let $\gamma > 0$ be the radius of the largest open sphere about q lying in D. According to Theorem 1 the integral expression

$$\int_{S_x} \phi(q + \rho x)\omega(\rho^{-1}x) = \rho^{-\mu} \int_{S_x} \phi(q + \rho x)\omega(x)$$

is independent of ρ for $0 < \rho < \gamma$. That the integral is a harmonic function of q in the open set D is easily seen by differentiating under the integral.

By letting ω be the constant harmonic function σ^{-1} homogeneous of degree zero, we obtain the mean value theorem; for the integral expression

$$\sigma^{-1} \int_{S_x} \phi(q + \rho x),$$

which is independent of ρ, has the limit $\phi(q)$ as $\rho \to 0$, whence

$$\phi(q) = \sigma^{-1} \int_{S_x} \phi(q + \rho x).$$

Thus the value of the harmonic function ϕ at the point q is the mean value of ϕ over the surface of any closed sphere centered at q and lying in D.

We next apply Theorem 1 to a pair of harmonic functions, both of which are homogeneous, to obtain a new proof of the following well-known theorem.

THEOREM 3. *Any two surface harmonics in E of different degrees are orthogonal.*

Proof. Let the surface harmonics be generated by the functions ω_1 and ω_2 harmonic in E and homogeneous about the origin of degrees μ_1 and μ_2 respectively. Then according to Theorem 1 the integral expression

$$\int_{S_x} \omega_1(\rho x)\omega_2(\rho^{-1}x) = \rho^{\mu_1-\mu_2} \int_{S_x} \omega_1(x)\omega_2(x)$$

is independent of ρ for $\rho > 0$. Since $\mu_1 \neq \mu_2$, we have

$$\int_{S_x} \omega_1(x)\omega_2(x) = 0.$$

Several characterization identities for harmonic functions may be obtained by suitable specialization of identity (1). The quadratic identity associated with (1) is an example of such a characterization identity.

THEOREM 4. *A function ϕ, continuous in an open set D of E, is harmonic in D if and only if the quadratic integral identity*

$$(4) \qquad \int_{S_x} \phi(q+\rho_1 x)\phi(q+\rho_2 x) = \int_{S_x} \phi(q+\rho x)^2, \qquad \rho = \sqrt{\rho_1\rho_2},$$

holds for all non-negative real numbers ρ_1 and ρ_2 such that the closed spheres of radii ρ_1 and ρ_2 about q lie in D.

Proof. If ϕ is harmonic in D then (4) is an immediate consequence of (1).

We now show that if (4) holds then ϕ is harmonic. We need consider only the case where $\rho_1 = 0$. Then $\rho = 0$ and the identity (4) may be written in the form

$$(5) \qquad \phi(q)[\phi(q) - \sigma^{-1} \int_{S_x} \phi(q+\rho_2 x)] = 0.$$

We shall say that ϕ has the *strong mean value property* at a point q of D if the value of ϕ at q is equal to the mean value of ϕ over the surfaces of *all* closed spheres about q lying in D; and we shall say that ϕ has the *weak mean value property* at q if the value of ϕ at q is equal to the mean value of ϕ over the surfaces of all sufficiently small closed spheres about q lying in D.

Let U be the set of points of D at which ϕ does not vanish and let V

be the interior of the set of points of D at which ϕ does vanish. Obviously ϕ has the weak mean value property at every point of V. Moreover, according to (5), ϕ has the strong mean value property at every point of U, and, since ϕ is continuous in D, it is easy to see that ϕ also has the strong mean value property at every point of $\bar{U}$ in D. Thus ϕ has the weak mean value property at every point of the set $\bar{U} + V$ in D. Finally, since $\bar{U}$ and V are complementary sets in D, the continuous function ϕ has the weak mean value property at every point of D and hence is harmonic in D.

We use the identity (4) to obtain a new proof of the following fundamental theorem. To my knowledge all previous proofs of this theorem have been founded upon series expansions for harmonic functions.[1]

THEOREM 5. *If the function ϕ is harmonic in a connected open set D of E and vanishes over some non-null open subset of D, then ϕ vanishes throughout D.*

Proof. Let V be the interior of the set of points of D at which ϕ vanishes. Thus V is an open subset of D; and, since ϕ vanishes over some non-null open subset of D, V is non-null. We shall show that V is closed in D.

Suppose, to the contrary, that V is not closed in D. Then there exists in D a boundary point p of V. Since D is open, some open sphere about p lies in D. Let the radius of such a sphere be $2\rho_2$. The point p is a boundary point of the open set V, so the open ρ_2 sphere about p contains a point q of V. Now the open ρ_2 sphere about q lies, together with its surface, in the open sphere of radius $2\rho_2$ about p and is therefore contained in D. Moreover this open ρ_2 sphere about q contains the boundary point p of V and hence does not lie in V. However, since q is a point of the open set V, some open sphere about q does lie in V. Let ρ_1 be the radius of the largest such open sphere about q lying in V; thus $0 < \rho_1 < \rho_2$. Now ϕ vanishes throughout the open ρ_1 sphere about q and, being continuous, vanishes on the surface of this sphere. Applying Theorem 4 we then obtain

$$\int_{S_x} \phi(q + \rho x)^2 = 0, \qquad \rho = \sqrt{\rho_1 \rho_2} > 0.$$

Consequently the function ϕ, being continuous, must vanish on the surface

[1] On page 226 of his book *Functions of a Complex Variable* (1936) W. F. Osgood says concerning this theorem: "It can not be proved by the integral representations, but requires the development into series."

of the open ρ sphere about q. Now ϕ is harmonic and hence vanishes throughout this open ρ sphere. But $\rho > \rho_1$ in contradiction to the definition of the open ρ_1 sphere about q.

This contradiction proves that V is closed in D. Thus V is a non-null set both open and closed in D. Since D is connected, we conclude that $V = D$. Therefore ϕ vanishes throughout D.

We use this result and the quadratic integral identity (4) to prove the following generalization of Liouville's theorem for harmonic functions.

THEOREM 6. *If the function ϕ is harmonic in E and vanishes at the origin, and if a constant μ exists such that*

$$\int_{S_x} \phi^+(\rho x) \leqq \mu, \qquad (\rho > 0), \tag{6}$$

where ϕ^+ is the continuous function defined at each point of E as ϕ if $\phi > 0$ and as 0 if $\phi \leqq 0$, then ϕ vanishes throughout E.

Proof. Let $\epsilon > 0$ be arbitrarily selected. Since ϕ is continuous and vanishes at the origin, a ρ can be chosen so large that

$$-\epsilon \leqq \phi(\rho^{-1}x) \leqq \epsilon \tag{7}$$

for all x in S. Moreover, since ϕ is harmonic and vanishes at the origin, we have by the mean value theorem

$$\int_{S_x} \phi(\rho x) = 0. \tag{8}$$

Using the above relations (8), (7), (6) we obtain the following inequality from the quadratic integral identity (4):

$$\int_{S_x} \phi(x)^2 = \int_{S_x} \phi(\rho^{-1}x)\phi(\rho x) = \int_{S_x} [\phi(\rho^{-1}x) + \epsilon]\phi(\rho x) \leqq 2\epsilon \int_{S_x} \phi^+(\rho x) \leqq 2\epsilon\mu.$$

This inequality holds for all $\epsilon > 0$, so that

$$\int_{S_x} \phi^+(\rho x)^2 = 0.$$

Consequently the function ϕ, being continuous, must vanish on the surface of the unit sphere. Now ϕ is harmonic and hence vanishes throughout the unit sphere. Therefore, according to Theorem 5, ϕ vanishes throughout E.

We note that any function ϕ which is bounded or merely bounded above satisfies condition (6).

As a final result we give another application of the quadratic integral identity (4). We shall say that a function ϕ defined on the surfaces of two concentric spheres of positive radii ρ_1 and ρ_2 about q changes sign between these two spheres on the ray from q determined by the unit vector x if

$$\phi(q+\rho_1 x)\phi(q+\rho_2 x) \leqq 0.$$

THEOREM 7. *If a function ϕ, harmonic in a connected open set D of E, changes sign between some two concentric closed spheres of positive radii lying in D on every ray from the common center of these two spheres, then ϕ vanishes throughout D.*

Proof. Let the two given concentric spheres have positive radii ρ_1 and ρ_2 and common center q. Since ϕ changes sign between these two spheres on every ray from q, we obtain from (4) the inequality

$$\int_{S_x} \phi(q+\rho x)^2 \leqq 0, \qquad \rho = \sqrt{\rho_1 \rho_2} > 0.$$

Consequently the function ϕ, being continuous, must vanish on the surface of the open ρ sphere about q and hence, being harmonic, must vanish throughout the connected open set D.

In conclusion we remark that if in the above theorems the function ϕ harmonic in D is continuous in the set $D+B$, where B is some subset of the boundary of D, then the closed spheres mentioned in the theorems may lie in $D+B$.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES.

THE LINEAR DIFFERENCE-DIFFERENTIAL EQUATION WITH ASYMPTOTICALLY CONSTANT COEFFICIENTS.*

By E. M. Wright.

1. Introduction. The general linear difference-differential equation is

$$\sum_{\mu=0}^{m} \sum_{\nu=0}^{n} A_{\mu\nu}(x) y^{(\nu)}(x + b_\mu) = v(x). \tag{1.1}$$

We suppose x a real variable,

$$m \geq 1, \qquad n \geq 1, \qquad y^{(0)}(x) \equiv y(x), \qquad 0 = b_0 < b_1 < \cdots < b_m,$$

and $v(x)$ and $A_{\mu\nu}(x)$ bounded and integrable in any finite interval. In [10] [1] I have discussed the existence, uniqueness and behaviour as $x \to \infty$ of solutions of the general equation (1.1) and in [11] I have added a little to the theory of the particular case of (1.1) in which the $A_{\mu\nu}(x)$ are all constants and $v(x)$ is zero, viz.

$$\Lambda(y) \equiv \sum_{\mu=0}^{m} \sum_{\nu=0}^{n} a_{\mu\nu} y^{(\nu)}(x + b_\mu) = 0. \tag{1.2}$$

The latter equation had already received much attention; [11] contains references to the literature, including accounts of applications to various practical problems. Any solution of (1.2) takes the form

$$y(x) = \sum_{r=1}^{\infty} P_r(x) e^{s_r x},$$

where the s_r are the zeros of

$$\tau(s) = \sum_{\mu=0}^{m} \sum_{\nu=0}^{n} a_{\mu\nu} s^\nu e^{b_\mu s}$$

and $P_r(x)$ is a polynomial of degree less than the order of the zero s_r of $\tau(s)$. I require here only the existence and order results for the general equation contained in [10] and a little information about the zeros of $\tau(s)$. None of the detailed results of [11] for the equation (1.2) is needed.

In the present paper I suppose that $A_{\mu\nu}(x) \to a_{\mu\nu}$ as $x \to +\infty$ and deduce results about the behaviour of the solutions of (1.1). These results

* Received April 26, 1947.

[1] Numbers in square brackets refer to the bibliography at the end of the paper.

are less precise than those known for (1.2) but very much more precise than those found for the general equation (1.1) in [10]. I am even able to find a certain pseudo-asymptotic relationship between solutions of (1.1) and those of

$$\Lambda(y) = v(x). \tag{1.3}$$

Apart from this last result, the theory is analogous to that developed by Poincaré [6], Perron [3], [4], [5] and Bochner [1] for differential and for difference equations. My method owes much to Bochner's though I use the theory of Fourier transforms where he is able to use a much simpler series device.

The theorems which I prove here find immediate application in the theory of non-linear difference-differential equations. In [9] I explained how a theory of small solutions of such equations could be built up in three stages. The third stage, which I dealt with in detail, enables us to find an asymptotic expansion of any solution which, together with its relevant derivatives, is $O(e^{-Cx})$ for some positive C as $x \to +\infty$. The second stage consists in proving, under suitable general conditions, that any solution which is $o(1)$ is in fact $O(e^{-Cx})$ for some positive C as $x \to +\infty$. I have found a variety of methods to prove this for particular equations, but the general result follows most readily from the results which I prove here. For example, if $y(x)$ is a solution of the equation

$$y'(x+1) = -\alpha y(x)\{1 + y(x+1)\} \qquad (\alpha > 0), \tag{1.4}$$

such that $y(x) \to 0$ as $x \to +\infty$, then $y(x)$ is also a solution of an equation

$$y'(x+1) = A(x)y(x),$$

where $A(x) \to -\alpha$ as $x \to +\infty$. It can be deduced from Theorem 2 and (2.1) of the present paper that $y(x) = O(e^{-Cx})$ for some $C > 0$ provided there is a strip $|\mathcal{R}(s)| < C + \epsilon$ in which

$$se^s + \alpha = 0$$

has no roots. (This is so unless $\alpha - \frac{1}{2}\pi$ is a multiple of 2π.) Hence $y'(x) = O(e^{-Cx})$ by (1.4), and the methods of [9] are applicable. I postpone a further discussion of these non-linear equations to another occasion.

2. Notation. In what follows x and t are real variables and $s = \sigma + it$ is a complex variable. The phrase "almost all x" means, as usual, "all x except for a set of measure zero." The various functions $y(x), y^{(\nu)}(x), v(x)$,

$A_{\mu\nu}(x), \cdots$ are frequently written $y, y^{(\nu)}, v, A_{\mu\nu}, \cdots$; wherever the variable is thus omitted it is x. The numbers μ and ν are always integers satisfying $0 \leq \mu \leq m$, $0 \leq \nu \leq n$. Except where a narrower restriction is explicitly stated, the phrase "for all ν such that $0 \leq \nu \leq n$" is to be understood in any conclusion or hypothesis involving ν; and similarly for μ. The symbols $\sum_\mu$, $\sum_\nu$ denote summation over these ranges. The positive numbers δ and ϵ are to be thought of as small; $x_0, x_1, x_2, \cdots$ are fixed positive numbers, the choice of x_2 being subsequent to that of δ. The numbers

$$B_1(\sigma),\ B_2(\sigma_3, \sigma_4),\ B_3(\delta),\ B_4(\sigma),\ B_5(\sigma_3, \sigma_4)$$

are positive functions of the parameters specified in each case, but otherwise depend only on the $a_{\mu\nu}$, b_μ and the bounds of the $A_{\mu\nu}$, being independent of x, t, y, $\bar{y}$, z, v, N and, except for $B_3(\delta)$, of δ. C is a positive number, not always the same at each occurrence, independent of δ, x, t, y, $\bar{y}$, z, v, N, but possibly depending on ϵ, $a_{\mu\nu}$, b_μ, σ_1, σ_2, σ_3, σ_4, σ, σ'.

We write

$$\Omega(y) \equiv \sum_\mu \sum_\nu (A_{\mu\nu} - a_{\mu\nu}) y^{(\nu)}(x + b_\mu),$$

so that (1.1) may be written

$$\Lambda(y) + \Omega(y) = v.$$

We suppose that at least one of the $a_{\mu n}$ is non-zero, that is $\sum_\mu | a_{\mu n} | \neq 0$, so that the differential order of (1.2) and (1.3) is in fact n.

$\mathfrak{M}$ is the set of the real parts of the zeros of $\tau(s)$ together with their limit points. If $a_{mn} = 0$, we suppose $\mathfrak{M}$ to contain also the "number" $+\infty$ while, if $a_{0n} = 0$, we suppose $\mathfrak{M}$ to contain the "number" $-\infty$. We discuss $\mathfrak{M}$ a little more fully later.

We say that a function $f(x)$ is $L^2(a, b)$ if f is of integrable square over the (finite or infinite) interval (a, b). Let f be $L^2(x_0, x)$ for some fixed x_0 and every finite $x > x_0$. We define $\omega(f)$ by the condition that $fe^{-\sigma x}$ is $L^2(x_0, +\infty)$ for all $\sigma > \omega(f)$ but not for any $\sigma < \omega(f)$. If $fe^{-\sigma x}$ is $L^2(x_0, +\infty)$ for all σ, we take $\omega(f) = -\infty$; if $fe^{-\sigma x}$ is not $L^2(x_0, +\infty)$ for any σ, we take $\omega(f) = +\infty$. Finally we write

$$\omega_n(y) = \max_\nu \omega(y^{(\nu)}).$$

Clearly the number $\omega(f)$ provides a measure of the behaviour of f as $x \to +\infty$ which is similar to the more familiar $\overline{\lim} (\log | f |)/x$. It may readily be shown that either f tends to a finite non-zero limit as $x \to +\infty$ or

$$\overline{\lim_{x\to+\infty}} \frac{\log |f|}{x} \leq \omega\left(\frac{df}{dx}\right).$$

From this we can deduce that

$$\overline{\lim_{x\to+\infty}} \frac{\log |y|}{x} \leq \omega_n(y). \tag{2.1}$$

We write

$$I_+(\sigma, f) = \int_0^\infty | fe^{-\sigma x} |^2 dx,$$

provided f exists for almost all $x > 0$ and the integral converges, and

$$I_-(\sigma, f) = \int_{-\infty}^0 | fe^{-\sigma x} |^2 dx$$

under the corresponding conditions. Again

$$I(\sigma, f) = I_+(\sigma, f) + I_-(\sigma, f),$$

if both the latter integrals converge. Where there is no ambiguity about σ, we write $I(f)$ for $I(\sigma, f)$ and so on. Using the well-known inequality

$$\left(\int_a^b | F_1F_2 | \, dx\right)^2 \leq \int_a^b | F_1 |^2 dx \int_a^b | F_2 |^2 dx,$$

we see that

$$\{I(f_1f_2)\}^2 \leq I(f_1^2)I(f_2^2) \tag{2.2}$$

and

$$\begin{aligned} I(\sum_{l=1}^L f_l) &= \int_{-\infty}^\infty | \sum_{l=1}^L f_l |^2 e^{-2\sigma x} dx \\ &= \sum_{l_1=1}^L \sum_{l_2=1}^L \int_{-\infty}^\infty | f_{l_1} f_{l_2} | e^{-2\sigma x} dx \\ &\leq \sum_{l_1=1}^L \sum_{l_2=1}^L \{I(f_{l_1})I(f_{l_2})\}^{\frac{1}{2}} \\ &= [\sum_{l=1}^L \{I(f_l)\}^{\frac{1}{2}}]^2 \end{aligned} \tag{2.3}$$

and so

$$I(\sum_{l=1}^L f_l) \leq L \sum_{l=1}^L I(f_l). \tag{2.4}$$

Similar inequalities hold for $I_+(f)$ and $I_-(f)$. Also if $\sigma' < \sigma''$,

$$I_+(\sigma', f) \geq I_+(\sigma'', f), \qquad I_-(\sigma', f) \leq I_-(\sigma'', f), \tag{2.5}$$

equality being possible only if $f = 0$ almost everywhere.

3. The principal results. We regard each of (1.1), (1.2) and (1.3) as an equation in the unknown function $y^{(n)}$ and the unknown numbers $y^{(\nu)}(x_0)$ $(\nu \leq n-1)$ for some x_0, the functions $y^{(\nu)}$ $(\nu \leq n-1)$ being defined by

$$y^{(\nu)}(x) = y^{(\nu)}(x_0) + \int_{x_0}^{x} y^{(\nu+1)}(\xi)\, d\xi.$$

If $y^{(n)}$ is a function of integrable square over every finite interval and if the values of $y^{(\nu)}(x_0)$ for $\nu \leq n-1$ are such that an equation is satisfied for almost all x, we say that y is a *solution* of the equation. A *solution* y for $x \geq x_0$ has the obvious corresponding meaning, the equation being satisfied for almost all $x \geq x_0$ and $y^{(n)}$ being $L^2(x_0, X)$ for every finite $X > x_0$. We remark that the existence of a solution for $x \geq x_0$ of (1.1), for example, implies that v is $L^2(x_0, X)$ for every finite $X > x_0$. Our present definition of a solution differs substantially from that of [10]; we return to this point in the proof of Lemma 7 of the present paper.

We say that two solutions y, z are *equivalent* (or *equivalent for* $x > x_0$) if $y^{(\nu)}(x_0) = z^{(\nu)}(x_0)$ $(\nu \leq n-1)$ for some x_0 and if $y^{(n)}(x) = z^{(n)}(x)$ for almost all x (or for almost all $x > x_0$). A solution of an equation, subject or not to additional conditions, is *effectively unique* if any other solution, satisfying the same additional conditions, is equivalent to the first.

We can now state our principal results in the form of theorems.

THEOREM 1. *If y is a solution of*

$$\Lambda(y) + \Omega(y) = v$$

for $x >$ some x_1, then either $\omega_n(y) = \omega(v)$ or $\omega_n(y)$ is greater than $\omega(v)$ and belongs to $\mathfrak{M}$.

THEOREM 2. *If y is a solution of*

$$\Lambda(y) + \Omega(y) = 0 \tag{3.1}$$

for $x >$ some x_1, then either $\omega_n(y)$ belongs to $\mathfrak{M}$ or y is effectively equivalent to the zero solution[2] *for $x >$ some x_4.*

Results about the relations between solutions of (1.1) and those of (1.3) can be deduced from the lemmas needed for the proof of Theorems 1 and 2 and are stated in **9**.

[2] That is, the solution in which $y = y^{(1)} = \cdots = y^{(n)} = 0$ for all x.

4. The zeros of $\tau(s)$ and the set $\mathfrak{M}$. In [11] I described the distribution of the zeros of $\tau(s)$ in the complex s-plane, deducing my results from those of Langer [2]. Here we need much less about $\tau(s)$ and its zeros and quote what we want from [11]. If $a_{mn} \neq 0$, $\mathfrak{M}$ is bounded above while, if $a_{0n} \neq 0$, $\mathfrak{M}$ is bounded below. If $a_{mn} = 0$, $\mathfrak{M}$ contains $+\infty$ by definition; its remaining members may or may not be bounded above. If $a_{mn} = 0$ and if at least one of the $a_{m\nu}$ is non-zero, $\mathfrak{M}$ (apart from $+\infty$) is not bounded above. If $a_{m\nu} = 0$ for all ν and $a_{m-1,n} \neq 0$, $\mathfrak{M}$ (apart from $+\infty$) is bounded above; this case was excluded by the hypotheses of [11] but the result is immediately deducible. When $a_{0n} = 0$, similar remarks hold good for boundedness below.

In every case there will be intervals (σ_1, σ_2) such that $\tau(\sigma + it) \neq 0$ for any real t and $\sigma_1 < \sigma < \sigma_2$. We always use σ_1, σ_2 for the end-points of such an interval (which may well be contained in a similar larger interval) and σ_3, σ_4 for numbers such that

$$\sigma_1 < \sigma_3 \leq \sigma_4 < \sigma_2.$$

In [11] I showed that

$$|\tau(s)| > C \max_{\mu,\nu} |a_{\mu\nu} s^\nu e^{b_\mu s}|,$$

whenever s is uniformly bounded from the zeros of $\tau(s)$. Since at least one of the $a_{\mu n}$ is non-zero, we have

LEMMA 1. *If $\sigma_1 + \epsilon \leq \sigma \leq \sigma_2 - \epsilon$, then*

$$|\tau(\sigma + it)| > C(1 + |t|^n). \tag{4.1}$$

In particular, (4.1) *is true for $\sigma_3 \leq \sigma \leq \sigma_4$.*

5. The equations with constant coefficients.

LEMMA 2. *If y is a solution of* (1.2) *such that $I_+(\sigma_4, y^{(\nu)})$ and $I_-(\sigma_3, y^{(\nu)})$ converge, then y is equivalent to the zero solution.*

Let $s = \sigma + it$ and $\sigma \geq \sigma_4 + \epsilon$. Since

$$\left(\int_0^X |y^{(\nu)} e^{-sx}| \, dx\right)^2 \leq I_+(\sigma_4, y^{(\nu)}) \int_0^X e^{-2\epsilon x} dx,$$

the integral

$$\int_0^\infty y^{(\nu)} e^{-sx} dx \tag{5.1}$$

is absolutely convergent for all ν. Integrating by parts we have, for $\nu \geq 1$,

$$\int_0^X y^{(\nu)} e^{-sx} dx = y^{(\nu-1)}(X) e^{-sX} - y^{(\nu-1)}(0) + s \int_0^X y^{(\nu-1)} e^{-sx} dx.$$

Let $X \to \infty$. Since (5.1) is absolutedy convergent for all ν, $y^{(\nu-1)}(X) e^{-sX}$ tends to a limit, which must be zero. Hence

$$\int_0^\infty y^{(\nu)} e^{-sx} dx = s \int_0^\infty y^{(\nu-1)} e^{-sx} dx - y^{(\nu-1)}(0)$$

and so

$$\int_0^\infty y^{(\nu)}(x+b) e^{-sx} dx = e^{bs} \int_0^\infty y^{(\nu)}(x) e^{-sx} dx - e^{bs} \int_0^b y^{(\nu)}(x) e^{-sx} dx$$

$$= e^{bs}\Big(s^\nu \int_0^\infty y(x) e^{-sx} dx - \sum_{l=1}^{\nu} s^{l-1} y^{(\nu-l)}(0) - \int_0^b y^{(\nu)}(x) e^{-sx} dx\Big).$$

Hence, by (1.2),

$$0 = \int_0^\infty \Lambda(y) e^{-sx} dx = \tau(s) \int_0^\infty y e^{-sx} dx - H(s), \tag{5.2}$$

where

$$H(s) = \sum_{\mu=0}^{m} \sum_{\nu=1}^{n} \sum_{l=1}^{\nu} a_{\mu\nu} s^{l-1} y^{(\nu-l)}(0) e^{b_\mu s}$$
$$+ \sum_{\mu=1}^{m} \sum_{\nu=0}^{n} a_{\mu\nu} e^{b_\mu s} \int_0^{b_\mu} y^{(\nu)} e^{-sx} dx.$$

Since $n \geq 1$, y is an integral and so is continuous and of bounded variation. Using (5.2) in the usual Laplace inversion formula, we have

$$\frac{1}{2\pi i} \int_{\sigma_4+\epsilon-i\infty}^{\sigma_4+\epsilon+i\infty} \frac{H(s)}{\tau(s)} e^{sx} ds = \begin{cases} y(x) & (x>0), \\ \tfrac{1}{2} y(x) & (x=0), \\ 0 & (x<0). \end{cases}$$

A similar argument leads to

$$-\frac{1}{2\pi i} \int_{\sigma_3-\epsilon-i\infty}^{\sigma_3-\epsilon+i\infty} \frac{H(s)}{\tau(s)} e^{sx} ds = \begin{cases} 0 & (x>0), \\ \tfrac{1}{2} y(x) & (x=0), \\ y(x) & (x<0), \end{cases}$$

and so, for all x,

$$y(x) = \frac{1}{2\pi i}\Big(\int_{\sigma_4+\epsilon-i\infty}^{\sigma_4+\epsilon+i\infty} - \int_{\sigma_3-\epsilon-i\infty}^{\sigma_3-\epsilon+i\infty}\Big) \frac{H(s) e^{sx}}{\tau(s)} ds. \tag{5.3}$$

Let us choose ϵ small enough to ensure that

$$\sigma_1 + \epsilon < \sigma_3 - \epsilon < \sigma_4 + \epsilon < \sigma_2 - \epsilon. \tag{5.4}$$

In the strip

$$\sigma_3 - \epsilon \leq \sigma \leq \sigma_4 + \epsilon,$$

the integral functions $\tau(s)$ and $H(s)$ satisfy

$$|\tau(s)| > C(1+|s|^n), \qquad |H(s)| < C(1+|s|^{n-1}),$$

the former inequality following from (5.4) and Lemma 1. Hence $H(s)/\tau(s)$ is regular and

$$\left|\frac{H(s)e^{sx}}{\tau(s)}\right| < \frac{Ce^{C|x|}}{|s|}$$

in the strip. It follows from (5.3) that $2\pi i y(x)$ is the limit as $T \to \infty$ of the integral of $e^{sx}H(s)/\tau(s)$ round a rectangle whose corners are at the points

$$\sigma_3 - \epsilon - iT, \qquad \sigma_4 + \epsilon - iT, \qquad \sigma_4 + \epsilon + iT, \qquad \sigma_3 - \epsilon + iT.$$

By Cauchy's theorem this integral is zero for all values of T and so $y = 0$ for all x. From this it follows that $y^{(\nu)} = 0$ for $\nu \leq n-1$ and all x and that $y^{(n)} = 0$ for almost all x.

Lemma 3. *If $\sigma_1 < \sigma < \sigma_2$ and if $I(\sigma, v)$ converges, there are a solution y of (1.3) and a number $B_1(\sigma)$ such that*

$$I(\sigma, y^{(\nu)}) \leq B_1(\sigma) I(\sigma, v).$$

We may suppose $\sigma = 0$ without loss of generality. For, if $\sigma \neq 0$, we make the transformations

$$v = e^{\sigma x}v_0, \qquad y = e^{\sigma x}y_0, \qquad \Lambda(y) = \Lambda(e^{\sigma x}y_0) = e^{\sigma x}\Lambda_0(y_0)$$

and observe that

$$y^{(\nu)}e^{-\sigma x} = \sum_{l=0}^{\nu} c_{l\nu}(\sigma) y_0^{(l)}, \qquad I(\sigma, v) = I(0, v_0),$$

and

$$I(\sigma, y^{(\nu)}) \leq \sum_{l=0}^{\nu} (\nu+1)|c_{l\nu}(\sigma)|^2 I(0, y_0^{(l)}),$$

where $c_{l\nu}(\sigma)$ depends only on l, ν and σ.

We use l. i. m. to denote "limit in mean square." By [7] (p. 69),

$$V(t) = \underset{X\to\infty}{\text{l. i. m.}} \int_{-X}^{X} v(x) e^{-ixt} dx$$

exists and is $L^2(-\infty, \infty)$ and so, by Lemma 1, $t^n V(t)/\tau(it)$ is $L^2(-\infty, \infty)$. We take

$$y_n(x) = \underset{T\to\infty}{\text{l. i. m.}} \frac{1}{2\pi} \int_{-T}^{T} \frac{(it)^n V(t) e^{ixt}}{\tau(it)} dt.$$

By [7] (p. 69), $y_n(x)$ exists and is almost everywhere equal to

$$\frac{1}{2\pi}\frac{d}{dx}\int_{-\infty}^{\infty}\frac{(it)^{n-1}V(t)}{\tau(it)}(e^{ixt}-1)dt$$
$$=\frac{1}{2\pi}\frac{d}{dx}\int_{-\infty}^{\infty}\frac{(it)^{n-1}V(t)e^{ixt}}{\tau(it)}dt.$$

If we take

$$y^{(\nu)}(x_0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{(it)^{\nu}V(t)e^{ix_0t}}{\tau(it)}dt \qquad (\nu\leq n-1)$$

for some x_0, it follows that

$$y^{(\nu)}(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{(it)^{\nu}V(t)e^{ixt}}{\tau(it)}dt \qquad (\nu\leq n-1),$$

both integrals being absolutely and uniformly convergent by Lemma 1. Hence

$$y^{(\nu)}(x)=\frac{1}{2\pi}\underset{T\to\infty}{\mathrm{l.\,i.\,m.}}\int_{-T}^{T}\frac{(it)^{\nu}V(t)e^{ixt}}{\tau(it)}dt$$

for all ν and

$$\Lambda(y)=\frac{1}{2\pi}\underset{T\to\infty}{\mathrm{l.\,i.\,m.}}\int_{-T}^{T}V(t)e^{ixt}dt=v(x)$$

for almost all x, by p. 69 of [7] and the definition of $V(t)$.

Finally, by the integral form of Parseval's theorem,

$$I(y^{(\nu)})=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{t^{2\nu}\,|\,V(t)\,|^{2}}{|\,\tau(it)\,|^{2}}dt$$
$$\leq\frac{B_1(0)}{2\pi}\int_{-\infty}^{\infty}|\,V(t)\,|^{2}dt=B_1(0)I(v).$$

LEMMA 4. *If $I_+(\sigma_4,v)$ and $I_-(\sigma_3,v)$ converge, there are a solution y of (1.3) and a number $B_2(\sigma_3,\sigma_4)$ such that*

$$(5.5)\qquad I_+(\sigma_4,y^{(\nu)})+I_-(\sigma_3,y^{(\nu)})\leq B_2(\sigma_3,\sigma_4)\{I_+(\sigma_4,v)+I_-(\sigma_3,v)\}.$$

If y_1 is any other solution of (1.3) such that $I_+(\sigma_4,y_1^{(\nu)})$ and $I_-(\sigma_3,y_1^{(\nu)})$ converge, then y and y_1 are equivalent.

We write

$$v_2=v,\qquad v_3=0\ \ (x\geq 0);\qquad v_2=0,\qquad v_3=v\ \ (x<0);$$

y_2 for the solution of $\Lambda(y)=v_2$ of Lemma 3 with $\sigma=\sigma_4$; and y_3 for the solution of $\Lambda(y)=v_3$ of Lemma 3 with $\sigma=\sigma_3$, the conditions of that lemma being clearly satisfied.

Now $y = y_2 + y_3$ is a solution of (1.3) satisfying (5.5). For

$$\Lambda(y) = \Lambda(y_2) + \Lambda(y_3) = v$$

for almost all x and, by (2.4), (2.5) and Lemma 3,

$$\begin{aligned} I_+(\sigma_4, y^{(\nu)}) &\leq 2I_+(\sigma_4, y_2^{(\nu)}) + 2I_+(\sigma_4, y_3^{(\nu)}) \\ &\leq 2I_+(\sigma_4, y_2^{(\nu)}) + 2I_+(\sigma_3, y_3^{(\nu)}) \\ &\leq 2I(\sigma_4, y_2^{(\nu)}) + 2I(\sigma_3, y_3^{(\nu)}) \\ &\leq 2B_1(\sigma_4) I(\sigma_4, v_2) + 2B_1(\sigma_3) I(\sigma_3, v_3) \\ &\leq \tfrac{1}{2} B_2(\sigma_3, \sigma_4) \{I_+(\sigma_4, v) + I_-(\sigma_3, v)\}. \end{aligned}$$

where $B_2(\sigma_3, \sigma_4) = 4B_1(\sigma_3) + 4B_1(\sigma_4)$. Similarly for $I_-(\sigma_3, y^{(\nu)})$.

If y_1 satisfies the hypothesis of the second part of Lemma 4, then $y - y_1$ is a solution of (1.2) satisfying the conditions of Lemma 2. Hence y and y_1 are equivalent and y is effectively unique.

6. Successive approximation. Let $\delta > 0$. Since $A_{\mu\nu} \to a_{\mu\nu}$ as $x \to +\infty$, we can choose a positive number $B_3(\delta)$ such that $| A_{\mu\nu} - a_{\mu\nu} | < \delta$ for all $x > B_3(\delta)$ and for every μ and ν. We now take $x_2 > B_3(\delta)$ and define the operator Δ by

$$\Delta(y) \equiv \sum_{\mu} \sum_{\nu} (A_{\mu\nu} - a_{\mu\nu}) y^{(\nu)}(x + b_\mu) \qquad (x \geq x_2),$$

$$\Delta(y) = 0 \qquad (x < x_2).$$

If $I_+(\sigma, y^{(\nu)})$ converges, we have by (2.4)

$$I_+\{\sigma, \Delta(y)\} \leq B_4(\sigma) \delta^2 \max_{\nu} \int_{x_2}^{\infty} | y^{(\nu)} e^{-\sigma x} |^2 \, dx \tag{6.1}$$

$$\leq B_4(\sigma) \delta^2 \max_{\nu} I(\sigma, y^{(\nu)}), \tag{6.2}$$

where $B_4(\sigma) = (m+1)(n+1)^2 \sum_{\mu} e^{2\sigma b_\mu}$. Trivially

$$I_-\{\sigma, \Delta(y)\} = 0 \tag{6.3}$$

for every σ.

LEMMA 5. *Let $I_+(\sigma_4, v)$ and $I_-(\sigma_3, v)$ converge,*

$$\delta \leq B_5(\sigma_3, \sigma_4) = \tfrac{1}{2} \{B_2(\sigma_3, \sigma_4) \, B_4(\sigma_4)\}^{-\frac{1}{2}}$$

and $x_2 > B_3(\delta)$. *Then the equation*

$$\Lambda(y) + \Delta(y) = v \tag{6.4}$$

has a solution y *for which*

$$I_+(\sigma_4, y^{(\nu)}) + I_-(\sigma_3, y^{(\nu)}) \leq 2B_2(\sigma_3, \sigma_4)\{I_+(\sigma_4, v) + I_-(\sigma_3, v)\}.$$

Any other solution $\bar{y}$ *of* (6.4), *for which* $I_+(\sigma_4, \bar{y}^{(\nu)})$ *and* $I_-(\sigma_3, \bar{y}^{(\nu)})$ *converge, is equivalent to* y.

We consider the sequence of equations

$$\Lambda(z_0) = v, \qquad \Lambda(z_N) = -\Delta(z_{N-1}) \qquad\qquad (N \geq 1),$$

with the condition that $I_+(\sigma_4, z_N^{(\nu)})$ and $I_-(\sigma_3, z_N^{(\nu)})$ are to converge for all $N \geq 0$. By Lemma 4, the first equation has an effectively unique solution z_0 and

$$I_+(\sigma_4, z_0^{(\nu)}) + I_-(\sigma_3, z_0^{(\nu)}) \leq B_2 D,$$

where

$$B_2 = B_2(\sigma_3, \sigma_4), \quad D = I_+(\sigma_4, v) + I_-(\sigma_3, v).$$

Hence, by (6.2) and (6.3),

$$I_+\{\sigma_4, \Delta(z_0)\} + I_-\{\sigma_3, \Delta(z_0)\} \leq B_2 B_4 \delta^2 D,$$

where $B_4 = B_4(\sigma_4)$.

Similarly the equation $\Lambda(z_1) = -\Delta(z_0)$ has a solution z_1 such that

$$I_+(\sigma_4, z_1^{(\nu)}) + I_-(\sigma_3, z_1^{(\nu)}) \leq B_2^2 B_4 \delta^2 D.$$

and

$$I_+\{\sigma_4, \Delta(z_1)\} + I_-\{\sigma_3, \Delta(z_1)\} \leq B_2^2 B_4^2 \delta^4 D.$$

Continuing the process, we see that our sequence of equations has a sequence of solutions $z_0, \ldots, z_N \ldots$ such that

$$I_+(\sigma_4, z_N^{(\nu)}) + I_-(\sigma_3, z_N^{(\nu)}) \leq B_2 D (B_2 B_4 \delta^2)^N \leq 2^{-2N} B_2 D,$$

since $B_2 B_4 \delta^2 \leq B_2 B_4 B_5^2 = \frac{1}{4}$. Hence, by (2.3),

$$I_+(\sigma_4, \sum_{M=N+1}^{N'} z_M^{(\nu)}) \leq [\sum_{M=N+1}^{N'} \{I_+(\sigma_4, z_M^{(\nu)})\}^{\frac{1}{2}}]^2$$

$$\leq B_2 D (\sum_{M=N+1}^{N'} 2^{-M})^2 \leq 2^{-2N} B_2 D \to 0$$

as $N \to \infty$ and similarly for $I_-(\sigma_3, \sum_{M=N+1}^{N'} z_M^{(\nu)})$. It follows by the usual theory of convergence in mean square (see, for example, [8], p. 386 *et seq.*) that, as $N \to \infty$, $\sum_{M=0}^{N} z_M^{(\nu)}$ converges in mean square over any finite interval to a function $y^{(\nu)}$. Also $y^{(\nu-1)}$ will be the integral of $y^{(\nu)}$ for almost all x and so, at our choice, for all x. Again $\sum_{M=0}^{N} z_M^{(\nu)} e^{-\sigma_4 x}$ converges in mean square over the interval $(0, +\infty)$ to $y^{(\nu)} e^{-\sigma_4 x}$ and there is a similar result over $(-\infty, 0)$. Finally

$$I_+(\sigma_4, y^{(\nu)}) + I_-(\sigma_3, y^{(\nu)}) \leq 4B_2 D.$$

Let us write

$$r_N = y - \sum_{M=0}^{N} z_M, \qquad h(x) = \Lambda(y) + \Delta(y) - v.$$

Since

$$\Lambda(\sum_{M=0}^{N} z_M) = v - \Delta(\sum_{M=0}^{N-1} z_M)$$

for almost all[3] x, we have

$$h(x) = \Lambda(r_N) + \Delta(r_{N-1})$$

for almost all x. Hence

$$\begin{aligned} I_+(\sigma_4, h) &+ I_-(\sigma_3, h) \\ &\leq C \max_{\nu} \{ I_+(\sigma_4, r_N^{(\nu)}) + I_+(\sigma_4, r_{N-1}^{(\nu)}) \\ &\qquad\qquad + I_-(\sigma_3, r_N^{(\nu)}) + I_-(\sigma_3, r_{N-1}^{(\nu)}) \} \\ &\to 0 \end{aligned}$$

as $N \to \infty$. But h is independent of N and so

$$I_+(\sigma_4, h) = 0 = I_-(\sigma_3, h).$$

Hence, almost everywhere, $h = 0$ and (6.4) is satisfied.

We have now to prove y effectively unique. Let $\bar{y}$ be another solution of (6.4) such that $I_+(\sigma_4, \bar{y})$ and $I_-(\sigma_3, \bar{y})$ converge. We solve the sequence of equations

$$\Lambda(\bar{r}_0) = -\Delta(\bar{y}), \qquad \Lambda(\bar{r}_N) = -\Delta(\bar{r}_{N-1}) \qquad (N \geq 1)$$

under the conditions that $I_4(\sigma_4, \bar{r}_N)$ and $I_-(\sigma_3, \bar{r}_N)$ converge and obtain (as before) a sequence of solutions $\{\bar{r}_N\}$ such that $I_+(\sigma_4, \bar{r}_N) \to 0$ and $I_-(\sigma_3, \bar{r}_N) \to 0$ as $N \to \infty$.

---

[3] By the definition of **3**, a *solution* need only satisfy an equation for *almost all x*.

If we now write

$$\bar{z}_0 = \bar{y} - \bar{r}_0, \qquad \bar{z}_N = \bar{r}_{N-1} - \bar{r}_N \qquad (N \geq 1),$$

we see that $I_+(\sigma_4, \bar{z}_N{}^{(\nu)})$ and $I_-(\sigma_3, \bar{z}_N{}^{(\nu)})$ converge for all $N \geq 0$. For almost all x

$$\Lambda(\bar{z}_0) = \Lambda(\bar{y}) - \Lambda(\bar{r}_0) = \Lambda(\bar{y}) + \Delta(\bar{y}) = v = \Lambda(z_0)$$

and so, by Lemma 4, z_0 and $\bar{z}_0$ are equivalent. Hence, for almost all x,

$$\Lambda(\bar{z}_1) = \Lambda(\bar{r}_0) - \Lambda(\bar{r}_1) = -\Delta(\bar{y}) + \Delta(\bar{r}_0) = -\Delta(\bar{z}_0)$$
$$= -\Delta(z_0)$$

and $\bar{z}_1$ and z_1 are equivalent. Continuing the argument, we see that corresponding members of the two sequences $\{z_N\}$ and $\{\bar{z}_N\}$ are equivalent. Hence, for almost all x,

$$\bar{y}^{(\nu)} - \sum_{M=0}^{N} z_M{}^{(\nu)} = \bar{y}^{(\nu)} - \sum_{M=0}^{N} \bar{z}_M{}^{(\nu)} = \bar{r}_N{}^{(\nu)}.$$

and so $\sum_{M=0}^{N} z_M{}^{(\nu)}$ converges in mean square over any finite interval to $\bar{y}^{(\nu)}$ as well as to $y^{(\nu)}$. It follows that y and $\bar{y}$ are equivalent.

7. Further lemmas.

LEMMA 6. *If* (i) y *is a solution of*

$$\Lambda(y) + \Omega(y) = v$$

for $x > x_1$, (ii) $\Delta(y)$ *is defined in terms of some* $x_2 \geq x$ *and* (iii) σ, σ' *are finite numbers such that* $ve^{-\sigma x}$ *and* $y^{(\nu)} e^{-\sigma' x}$ *are* $L^2(x_2, +\infty)$, *then there are a* $\bar{y}$ *and a* $\bar{v}$ *such that*

$$\Lambda(\bar{y}) + \Delta(\bar{y}) = \bar{v} \tag{7.1}$$

for almost all x,

$$\bar{y}^{(\nu)} = y^{(\nu)} \quad (x > x_2), \qquad \bar{y}^{(\nu)} = 0 \quad (x < x_2 - b_m) \tag{7.2}$$

$$\bar{v} = v \quad (x > x_2), \qquad \bar{v} = 0 \quad (x < x_2 - 2b_m) \tag{7.3}$$

and

$$I_+(\sigma, \bar{v}), \quad I_+(\sigma', \bar{y}^{(\nu)}), \quad I_-(\bar{\sigma}, \bar{v}), \quad I_-(\bar{\sigma}, \bar{y}^{(\nu)})$$

all converge, the last two for every finite $\bar{\sigma}$.

We define the numbers $a_0, \cdots, a_{n-1}$ in succession by the relation

$$(-1)^L a_L = \sum_{l=0}^{L-1} (-1)^{l+1} \binom{n}{L-l} a_l + \frac{(-1)^{n-1}}{L!} \sum_{l=L}^{n-1} \frac{(-1)^l b_m{}^l y^{(l)}(x_2)}{(l-L!)}.$$

In the interval $x_2 - b_m \leq x \leq x_2$ we put

$$\bar{y} = \sum_{\nu=0}^{n-1} \frac{y^{(\nu)}(x_2)}{\nu!} (x-x_2)^\nu + \frac{(x-x_2)^n}{b_m{}^n} \sum_{l=0}^{n-1} \frac{a_l(x-x_2+b_m)^l}{b_m{}^l}$$

and define $\bar{y}^{(\nu)}$ by differentiation. Outside this interval $\bar{y}^{(\nu)}$ is defined by (7. 2). It can then be readily verified that $\bar{y}^{(\nu-1)}$ is continuous at $x = x_2 - b_m$ and at $x = x_2$ and so is the integral of $\bar{y}^{(\nu)}$ for all x. Finally we define $\bar{v}$ by (7. 1), so that (7. 3) is an immediate consequence of (7. 2). The last part of the lemma is then trivial.

LEMMA 7. *If, for $x \geq x_3$,* (i) $|A_{mn}| > C$, (ii) $|A_{\mu\nu}| < C$, (iii) *y is a solution of* (1. 1) *and* (iv) $\int_{x_3}^{X} |v|^2 dx < Ce^{CX}$ *for all $X > x_3$, then*

$$\int_{x_3}^{X} |y^{(n)}|^2 dx < Ce^{CX}$$

for all $X > x_3$, and $|y^{(\nu)}| < Ce^{Cx}$ for $\nu \leq n-1$ and all $x \geq x_3$.

This lemma is Theorem 3 (ii) of [10], apart from two differences in the definition of a *solution.* In [10] y is said to be a solution of (1. 1) for $x \geq x_3$ if $y^{(n)}$ satisfies (1. 1) for *all* $x \geq x_3$ and is integrable over every finite interval (x_3, X); in the present paper $y^{(n)}$ has to satisfy (1. 1) for *almost all* $x \geq x_3$ but must be of integrable square over every finite (x_3, X). Since $|A_{mn}| > C$, the first difference can be overcome by changing the value of the $y^{(n)}(x+b_m)$ of the present paper at every x of a set of measure zero so that (1. 1) is satisfied for all $x \geq x_3$; this does not alter the value of $y^{(\nu)}$ for $\nu \leq n-1$ nor that of any integral of $|y^{(n)}|^2$. The second difference is in the opposite direction; the $y^{(n)}$ of the present paper satisfies the stronger condition and, in particular, we can omit the condition in Theorem 3 (ii) of [10] that $y^{(n)}$ is of integrable square over the interval $(x_3, x_3 + b_m)$, since this is now so by definition.

LEMMA 8. *If $a_{mn} \neq 0$, if $\omega(v) < +\infty$ and if y is a solution of* (1. 1) *for $x \geq x_1$, then $\omega_n(y) < +\infty$.*

Since $a_{mn} \neq 0$ and $A_{\mu\nu} \to a_{\mu\nu}$ as $x \to +\infty$, there is a number $x_3 \geq x_1$

such that $|A_{mn}| > C$ and $|A_{\mu\nu}| < C$ for $x \geq x_3$. Since $\omega(v) < +\infty$, we can choose a positive $\sigma > \omega(v)$ such that $ve^{-\sigma x}$ is $L^2(x_3, +\infty)$. Hence

$$\int_{x_3}^{X} |v|^2 dx \leq e^{2\sigma X} \int_{x_3}^{\infty} |v|^2 e^{-2\sigma x} dx \leq Ce^{2\sigma X}.$$

Thus all the conditions of Lemma 7 are satisfied. Hence, for $\nu \leq n-1$ and some positive α, $y^{(\nu)} e^{-\alpha x}$ is $L^2(x_3, +\infty)$. The same is true for $\nu = n$ since, on integration by parts,

$$\int_{x_3}^{X} |y^{(n)} e^{-\alpha x}|^2 dx = e^{-2\alpha X} \int_{x_3}^{X} |y^{(n)}|^2 dx + 2\alpha \int_{x_3}^{X} \int_{x_3}^{x} |y^{(n)}(\xi)|^2 d\xi \,.\, e^{-2\alpha x} dx$$

$$\leq Ce^{(C-2\alpha)X} + C\alpha \int_{x_3}^{X} e^{(C-2\alpha)x} dx,$$

where the C are independent of α. The last expression is certainly bounded for suitable α. Hence $\omega_n(y) < +\infty$.

LEMMA 9. *If, for $x \geq x_4$,* (i) $|A_{0n}| > C$, (ii) $|A_{\mu\nu}| < C$, (iii) *y is a solution of* (3.1), (iv) $\int_{x_4}^{\infty} |y^{(n)}| e^{kx} dx$ *converges for every finite k and* (v) $|y^{(0)}|$ *does not tend to infinity nor to any non-zero limit as $x \to +\infty$, then y is equivalent to the zero solution for $x \geq x_4$.*

This is Theorem 4 of [10]. Since $|A_{0n}| > C$, we can modify $y^{(n)}(x)$ in (3.1) at a set of measure zero so as to satisfy (3.1) for *all* $x \geq x_4$.

LEMMA 10. *If $a_{0n} \neq 0$, if $\omega_n(y) = -\infty$ and if y is a solution of* (3.1) *for $x \geq x_1$, then y is equivalent to the zero solution for $x \geq$ some x_4.*

Since $a_{0n} \neq 0$, there is an $x_4 \geq x_1$ such that $|A_{0n}| > C$ and $|A_{\mu\nu}| < C$ for all $x \geq x_4$. Since $\omega_n(y) = -\infty$, $y^{(\nu)} e^{-\sigma x}$ is $L^2(x_4, +\infty)$ for every finite σ. Hence $|y^{(0)}|$ cannot tend to infinity nor to any non-zero limit as $x \to +\infty$. Also, for any $\epsilon > 0$,

$$\left(\int_{x_4}^{X} |y^{(n)}| e^{kx} dx\right)^2 \leq \int_{x_4}^{X} |y^{(n)}|^2 e^{2(k+\epsilon)x} dx \int_{x_4}^{X} e^{-2\epsilon x} dx < C$$

for all X and so condition (iv) of Lemma 9 is satisfied. Our result follows at once from that lemma.

8. Proof of Theorems 1 and 2. It follows trivially from (1.1) and the boundedness of $A_{\mu\nu}$ that $\omega_n(y) \geq \omega(v)$. Let us suppose that Theorem 1 is false, so that $\omega_n(y) > \omega(v)$ and $\omega_n(y)$ does not belong to $\mathfrak{M}$. If this is so it is impossible that $\omega_n(y) = +\infty$; for in that case $a_{mn} \neq 0$ by the

definition of $\mathfrak{M}$, $\omega(v) < \omega_n(y) = +\infty$ and, by Lemma 8, $\omega_n(y) < +\infty$, a contradiction. Hence $\omega_n(y)$ is finite, is greater than $\omega(v)$ and does not belong to the closed set $\mathfrak{M}$. We can therefore find numbers σ_1, σ_2, σ'_3, σ'_4 such that

$$\omega(v) < \sigma_1 < \sigma'_3 < \omega_n(y) < \sigma'_4 < \sigma_2 \tag{8.1}$$

and σ_1, σ_2 have their usual property that no σ of $\mathfrak{M}$ satisfies $\sigma_1 < \sigma < \sigma_2$, so that $\tau(\sigma + it) \neq 0$ in this interval.

We take

$$\delta \leq \min\{B_5(\sigma'_3, \sigma'_4),\ B_5(\sigma'_3, \sigma'_3)\},\ \ x_2 > \max\{x_1, B_3(\delta)\}$$

and define Δ in terms of x_2 as in **6**. By (8.1), $e^{-\sigma'_3 x}$ and $y^{(\nu)}e^{-\sigma'_4 x}$ and $L^2(x_2, +\infty)$. Hence, by Lemma 6, we have a $\bar{y}$ and a $\bar{v}$ such that

$$\Lambda(\bar{y}) + \dot{\Delta}(\bar{y}) = \bar{v} \tag{8.2}$$

for almost all x, $\bar{y}^{(\nu)} = y^{(\nu)}$ for $x > x_2$ and

$$I_+(\sigma'_3, \bar{v}),\ I_+(\sigma'_4, \bar{y}^{(\nu)}),\ I_-(\sigma'_3, \bar{v}),\ I_-(\sigma'_3, \bar{y}^{(\nu)})$$

all converge. By Lemma 5, with $\sigma_3 = \sigma'_3$, $\sigma_4 = \sigma'_4$, $\bar{y}$ is effectively unique for a given $\bar{v}$. By the same lemma with $\sigma_3 = \sigma_4 = \sigma'_3$, since $I_+(\sigma'_3, \bar{v})$ and $I_-(\sigma'_3, \bar{v})$ converge, (8.2) has an effectively unique solution $\bar{y}_1$ such that $I_+(\sigma'_3, \bar{y}_1^{(\nu)})$ and $I_-(\sigma'_3, \bar{y}_1^{(\nu)})$ converge and so, *a fortiori*, $I_+(\sigma'_4, \bar{y}_1^{(\nu)})$ converges. Hence $\bar{y}$ and $\bar{y}_1$ are equivalent and $I_+(\sigma'_3, \bar{y}^{(\nu)})$ converges. Since $y^{(\nu)} = \bar{y}^{(\nu)}$ for $x > x_2$, it follows that $\omega_n(y) \leq \sigma'_3$, which contradicts (8.1). Hence Theorem 1 is true.

To prove Theorem 2, we put $v = 0$ in Theorem 1, so that (1.1) becomes (3.1) and $\omega(v) = -\infty$. Hence $\omega_n(y)$ belongs to $\mathfrak{M}$ unless $\omega_n(y) = -\infty$ and $-\infty$ does not belong to $\mathfrak{M}$. In this exceptional case, however, $a_{0n} \neq 0$ and, by Lemma 10, y is equivalent to the zero solution for $x >$ some x_4. This is Theorem 2.

9. Further results. We can deduce from our lemmas a connection between solutions of (1.1) and those of (1.3).

THEOREM 3. *If* (i) *y is a solution of* (1.1) *for $x > x_1$,* (ii) *σ is a finite number not belonging to $\mathfrak{M}$,* (iii) *$\int_{x_1}^{\infty} |y^{(\nu)}|^2 e^{-2\sigma x}dx$ converges and* (iv) *η is any positive number, then there are a number $x_2 = x_2(\eta) \geq x_1$ and a solution z of* (1.3) *for $x \geq x_2$ such that*

$$(9.1)\qquad \max_\nu \int_{x_2}^{\infty} | y^{(\nu)} - z^{(\nu)} |^2 e^{-2\sigma x} dx < \eta \max_\nu . \int_{x_2}^{\infty} | y^{(\nu)} |^2 e^{-2\sigma x} dx$$

and

$$(9.2)\qquad \max_\nu \int_{x_2}^{\infty} | y^{(\nu)} - z^{(\nu)} |^2 e^{-2\sigma x} dx < \eta \max_\nu \int_{x_2}^{\infty} | z^{(\nu)} |^2 e^{-2\sigma x} dx.$$

The proof is not lengthy. For any $\delta > 0$, we can find a number $x_2 \geq \max\{x_1, B_3(\delta)\}$ and define $\Delta(y)$ in terms of x_2 so that

$$(9.3)\qquad I(\sigma, \Delta(y)) \leq B_4(\sigma)\delta^2 \max_\nu \int_{x_2}^{\infty} | y^{(\nu)} |^2 e^{-2\sigma x} dx$$

by (6.1). By hypotheses (i) and (iii), $ve^{-\sigma x}$ is $L^2(x_1, +\infty)$. Hence, by Lemma 6 with $\bar{\sigma} = \sigma' = \sigma$, there are a $\bar{y}$ and a $\bar{v}$ such that $\bar{y}^{(\nu)} = y^{(\nu)}$ and $\bar{v} = v$ for $x > x_2$, $I(\sigma, \bar{v})$ and $I(\sigma, \bar{y})$ converge, and

$$(9.4)\qquad \Lambda(\bar{y}) + \Delta(\bar{y}) = \bar{v}$$

for almost all x. Hence, by Lemma 4, with $\sigma_3 = \sigma_4 = \sigma$,

$$(9.5)\qquad \Lambda(z) = \bar{v}$$

has an effectively unique solution such that $I(\sigma, z^{(\nu)})$ converges. From (9.4) and (9.5),

$$\Lambda(\bar{y} - z) = -\Delta(\bar{y})$$

for almost all x and so

$$I(\sigma, \bar{y}^{(\nu)} - z^{(\nu)}) \leq B_2(\sigma, \sigma) I(\sigma, \Delta(\bar{y}))$$

by Lemma 4. But $\bar{y}^{(\nu)} = y^{(\nu)}$ for $x > x_2$ and so

$$\begin{aligned} \int_{x_2}^{\infty} | y^{(\nu)} - z^{(\nu)} |^2 e^{-2\sigma x} dx &= \int_{x_2}^{\infty} | \bar{y}^{(\nu)} - z^{(\nu)} |^2 e^{-2\sigma x} dx \\ &\leq I(\sigma, \bar{y}^{(\nu)} - z^{(\nu)}) \leq B_2(\sigma, \sigma) I(\sigma, \Delta(\bar{y})) \\ &\leq B_4(\sigma) B_2(\sigma, \sigma) \delta^2 \max_\nu \int_{x_2}^{\infty} | y^{(\nu)} |^2 e^{-2\sigma x} dx \end{aligned}$$

by (9.3). Since B_2 and B_4 are independent of δ, we can choose δ so that (9.1) and (9.2) follow at once.

THEOREM 4. *If* (i) z *is a solution of* (1.3) *for* $x > x_1$, (ii) σ *is a finite number not belonging to* $\mathfrak{M}$, (iii) $\int_{x_1}^{\infty} | z^{(\nu)} |^2 e^{-2\sigma x} dx$ *converges and* (iv) η *is any positive number, then there are a number* $x_2 = x_2(\eta) \geq x_1$ *and a solution* y *of* (1.1) *for* $x \geq x_2$ *such that* (9.1) *and* (9.2) *are true.*

The proof of this is a little longer than that of Theorem 3 and involves the construction of the y of Lemma 5, putting $z_0^{(\nu)} = z^{(\nu)}$ $(x > x_2)$, $z_0^{(\nu)} = 0$ $(x < x_2 - b_m)$ and defining suitable $z_0^{(\nu)}$ for $x_2 \leq x \leq x_2 - b_m$ to secure the necessary continuity of $z_0^{(\nu)}$ for $\nu \leq n - 1$ much as in Lemma 6. The details are sufficiently obvious and I omit them.

UNIVERSITY OF ABERDEEN,
SCOTLAND.

BIBLIOGRAPHY

[1]. S. Bochner, "Allgemeine lineare Differenzengleichungen mit asymptotisch konstanten Koeffizienten," *Mathematische Zeitschrift*, vol. 33 (1931), pp. 426-450.

[2]. R. E. Langer, "The asymptotic location of the roots of a certain transcendental equation," *Transactions of the American Mathematical Society*, vol. 31 (1929), pp. 837-844.

[3]. O. Perron, "Über Systeme von linearen Differenzengleichungen erster Ordnung," *Journal für Mathematik*, vol. 147 (1917), pp. 36-53.

[4]. O. Perron, "Über Summengleichungen und Poincarésche Differenzengleichungen," *Mathematische Annalen*, vol. 84 (1921), pp. 1-15.

[5]. O. Perron, "Über Stabilität und asymptotisches Verhalten der Lösungen eines Systems endlicher Differenzengleichungen," *Journal für Mathematik*, vol. 161 (1929), pp. 41-64.

[6]. H. Poincaré, "Sur les équations linéaires aux différentielles ordinaires et aux différences finies," *American Journal of Mathematics*, vol. 7 (1885), pp. 203-258.

[7]. E. C. Titchmarsh, *Theory of Fourier Integrals*, Oxford (1937).

[8]. E. C. Titchmarsh, *Theory of Functions*, Oxford (1932).

[9]. E. M. Wright, "The non-linear difference-differential equation," *Quarterly Journal of Mathematics* (Oxford series), vol. 17 (1946), pp. 245-252.

[10]. E. M. Wright, "Linear difference-differential equations," *Proceedings of the Cambridge Philosophical Society* (in press).

[11]. E. M. Wright, "The linear difference-differential equation with constant coefficients," *Proceedings of the Royal Society of Edinburgh* (in press).

SUFFICIENT CONDITIONS FOR MULTIPLE INTEGRAL PROBLEMS IN THE CALCULUS OF VARIATIONS.*

By Magnus R. Hestenes.

1. Introduction. The problem at hand is that of minimizing an integral of the form

$$I(y) = \int_S f(x, y, \dot{y})\, dx$$

in a class of functions

$$y(x) = y(x_1, \cdots, x_m) \qquad (x \text{ on } S)$$

in $(x_1, \cdots, x_m, y)$-space having prescribed values on the boundary of S and satisfying a set of isoperimetric conditions

$$(1.1) \qquad I_\sigma(y) = \int_S f_\sigma(x, y, \dot{y})\, dx = k_\sigma \qquad (\sigma = 1, \cdots, r).$$

Here $\dot{y}$ denotes the set $(\partial y/\partial x_1, \cdots, \partial y/\partial x_m)$.

The purpose of the present paper is to establish a set of sufficient conditions for this problem. The set here established has the same generality as the corresponding theorem for simple integrals (See 8, 9).[1] There are essentially three methods of establishing sufficiency theorems in the calculus of variations. The first is by constructing a Mayer field. This method has been very effective for simple integrals. An excellent account of the theory of Mayer fields for multiple integrals has been given by Bliss (2). As has been pointed out by Bliss (3) the existence of Mayer fields for multiple integral problems is difficult to establish. In fact except for the obvious case in which y does not appear explicitly in the integrand, the only published account in which the existence of Mayer fields has been established for multiple integrals appears to be the one by Lichtenstein (13) in 1917. Moreover this method is not applicable in a natural way to isoperimetric problems. The second method is by the expansion methods developed by Levi (12) and Reid (18) for simple integral problems. Reid (19) has announced, but not yet published, a sufficiency theorem for a strong relative minimum for multiple

* Received June 25, 1947.

[1] Roman numerals in parentheses refer to the list of references at the end of the paper.

integral problems without isoperimetric conditions, based upon expansion methods. The third method is the indirect method developed for simple integrals by McShane (14), Myers (17), and Hestenes (8, 9). This method has been extended by Karush (11) to multiple integral problems without isoperimetric conditions so as to obtain sufficient conditions for a semi strong relative minimum. In the present paper we extend the indirect method so as to obtain a strong sufficiency theorem for the problem here described. The case in which there is more than one dependent variable y but no isoperimetric side conditions is being treated by H. L. Meyer. Free use has been made of the ideas of Calkin (4) and Morrey (15, 16) on absolutely continuous functions.

One of the chief features of the indirect method is that it has been successful where other methods apparently have failed. It is particularly adapted to the study of isoperimetric problems in the sense that the introduction of isoperimetric side conditions does not significantly alter the arguments in a sufficiency proof. This is not the case when the other two methods described above are used. As a matter of fact these methods do not appear to be applicable to isoperimetric multiple integral problems.

The present paper is incomplete in the sense that a further study should be made of conditions which imply that the second variation is positive definite. It is clear that if there exists an extremal for the second variation that does not vanish on the closure of S, then the second variation is positive definite. A study of the second variation will be made in a later paper.

2. Preliminary remarks. Throughout the present paper the symbol x will denote a point $(x_1, \cdots, x_m)$ in an m-dimensional euclidean space. The derivative of a function $y(x)$ with respect to x_k will be denoted by either of the symbols $\dot{y}_k$, $\partial y/\partial x_k$. We shall use the notation $\dot{y}$ for the vector $(\dot{y}_1, \cdots, \dot{y}_m)$ determined by these derivatives. Given a vector $p = (p_1, \cdots, p_m)$ we distinguish between the notations $|p_k|$ and $|p|$. The first is the absolute value of the k-th coordinate of p, the second is the length of p, that is $|p| = (p_1{}^2 + \cdots + p_m{}^2)^{\frac{1}{2}}$. A repeated index in a term will denote summation with respect to that index over its range. The range usually will be clear from the context and will be omitted frequently. We make two exceptions to this rule. The indices k and q will never be summed. They have the ranges $1, 2, \cdots, m$ and $1, 2, 3, \cdots$ respectively.

When we wish to fix our attention on a particular coordinate x_k we write (x_k, x'_k) for x. Here the symbol x'_k denotes the point in the euclidean $(m-1)$-space determined by the remaining $m-1$ coordinates of x. The

interval $a_i \leqq x_i \leqq b_i$ $(i = 1, \cdots, m)$ will be designated by $[a, b]$ and $[a_k, b_k]$, $[a'_k, b'_k]$ will signify its projections in x_k-space and x'_k-space respectively. We shall use the abbreviations

$$\int_a^b y(x)\, dx = \int_{a_1}^{b_1} \cdots \int_{a_m}^{b_m} y(x)\, dx_1 \cdots dx_m$$

$$\int_{a'_k}^{b'_k} y(x_k, x'_k)\, dx'_k = \int_{a_1}^{b_1} \cdots \int_{a_{k-1}}^{b_{k-1}} \int_{a_{k+1}}^{b_{k+1}} \cdots \int_{a_m}^{b_m} y(x)\, dx_1 \cdots dx_{k-1} dx_{k+1} \cdots dx_m,$$

$$\int_{a'_k}^{b'_k} y \Big|_{a_k}^{b_k} dx'_k = \int_{a'_k}^{b'_k} \{y(b_k, x'_k) - y(a_k, x'_k)\}\, dx'_k.$$

The function

$$(2.1) \qquad y^h(x) = (2h)^{-m} \int_{-h}^{h} y(x + t)\, dt$$
$$= (2h)^{-m} \int_{-h}^{h} \cdots \int_{-h}^{h} y(x_1 + t_1, \cdots, x_m + t_m)\, dt_1 \cdots dt_m$$

will be called the integral mean of y. Here h is a positive number. We shall have occasion to use the relations

$$(2.2) \qquad \lim_{h=0} \int_S |\, y^h - y \,|^{\alpha}\, dx = 0 \qquad (\alpha \geqq 1)$$

$$(2.3) \qquad \int_S |\, y^h |^{\alpha}\, dx \leqq \int_S |\, y \,|^{\alpha}\, dx \qquad (\alpha \geqq 1)$$

which hold in case y is in the Lebesgue class $\mathfrak{L}_1$ on a neighborhood of S and $y = 0$ exterior to S. In case $y(x)$ is absolutely continuous in x_k for almost all x'_k and its derivative $\dot{y}_k$ is integrable, we have

$$(2.4) \qquad \dot{y}_k{}^h = \partial y^h / \partial x_k = (2h)^{-m} \int_{-h}^{h} \dot{y}_k(x + t)\, dt.$$

It will be assumed that the reader is familiar with the properties of the integral mean.

Let $\mathfrak{R}$ be an open set of points $(x_1, \cdots, x_m, y, p_1, \cdots, p_m)$. A function $f(x, y, p)$ will be called *admissible* if it is continuous on $\mathfrak{R}$ and has continuous derivatives on $\mathfrak{R}$ of the first and second orders with respect to the variables $y, p_1, \cdots, p_m$. It will be assumed throughout that the functions $f(x, y, p)$ and $f_\sigma(x, y, p)$ appearing in (1.1) and (1.2) are admissible.

We center our attention upon a particular function

$$y_0(x) \qquad x \text{ on } S$$

whose minimizing properties we shall study. The set S is the closure of a bounded open set whose boundary is of measure zero. The function $y_0(x)$ is assumed to be continuous on the closure of S and to have continuous derivatives $\dot{y}_{0k}(x)$ interior to S with continuous limits on the boundary of S. The elements $(x, y_0(x), \dot{y}_0(x))$ are assumed to be in $\mathfrak{R}$ for each point x on S. It is understood that $y_0(x)$ satisfies the isoperimetric conditions (1.2). It will be convenient to assume that the functions $y_0(x)$ and $\dot{y}_{0k}(x)$ have been extended so as to be continuous for all values of x. At points exterior to S the function $\dot{y}_{0k}(x)$ will not in general represent the derivative of $y_0(x)$ with respect to x_k.

In the next section we shall impose further restrictions on the function y_0 which will imply that y_0 minimizes $I(y)$ on a certain class of functions. For the purpose of sufficiency theorems it is desirable to select this class to be as large as conveniently possible. In fact we shall admit functions which are discontinuous on a set of measure zero. Our functions will be defined for all values of x, even though we are primarily interested in their values on the set S described in the last paragraph. The specific properties a comparison function $y(x)$ is required to possess are the following

(a) The function $y(x)$ is bounded and integrable on S.

(b) The difference function $\eta(x) = y(x) - y_0(x)$ when extended to be zero on the complement of S is absolutely continuous in each variable x_k for almost all x'_k.

(c) The derivatives $\dot{y}_k(x) = \partial y/\partial x_k$ $(k = 1, \cdots, m)$ are integrable on S. At points at which $\dot{y}_k(x)$ fails to exist set $\dot{y}_k(x) = \dot{y}_{0k}(x)$.

(d) The elements $(x, y(x), \dot{y}(x))$ are in $\mathfrak{R}$ for almost all points x on S.

(e) The integrals $I(y)$, $I_\sigma(y)$ have well defined values, finite or infinite.

The class of functions $y(x)$ having these properties will be denoted by $\mathfrak{A}$. If a function $y(x)$ in $\mathfrak{A}$ is continuous it is absolutely continuous on S in the sense of Tonelli, i. e. (1), $y(x)$ is continuous on S, (2), given a closed interval $[a, b]$ interior to S, the function $y(x)$ is absolutely continuous in each component x_k on $[a_k, b_k]$ for almost all x'_k on $[a'_k, b'_k]$, and (3), the derivatives $\dot{y}_k(x)$ are integrable on S. Conversely, every function $y(x)$ that is absolutely continuous on S in the sense of Tonelli and that coincides with $y_0(x)$ on the boundary of S belongs to $\mathfrak{A}$.

We shall be interested particularly in the subclass $\mathfrak{A}_0$ of $\mathfrak{A}$ on which

the relation $I_\sigma(y) = I_\sigma(y_0)$ $(\sigma = 1, \cdots, r)$ holds. We seek conditions on y_0 that will imply the existence of a neighborhood $\mathfrak{F}$ of y_0 in xy-space such that the inequality $I(y) \geqq I(y_0)$ holds for every function $y(x)$ in $\mathfrak{A}_0$ lying in $\mathfrak{F}$. A function $y(x)$ will be said to lie in $\mathfrak{F}$ if its elements $(x, y(x))$ are in $\mathfrak{F}$ for almost all x in S.

We shall be interested in three classes of *variations,* denoted respectively by $\mathfrak{B}'$, $\mathfrak{B}''$, $\mathfrak{B}_0$. A function $\eta(x)$ in $\mathfrak{B}'$ is characterized by the following three properties:

(i) The function $\eta(x)$ is integrable on S and is identically zero on the complement of S.

(ii) The function $\eta(x) = \eta(x_k, x'_k)$ is absolutely continuous in each component x_k for almost all x'_k $(k = 1, \cdots, m)$.

(iii) The derivatives $\dot\eta_k(x)$ $(k = 1, \cdots, m)$ of $\eta(x)$ with respect to x_k are integrable on S.

The subclass of $\mathfrak{B}'$ consisting of all variations $\eta(x)$ having integrable square derivatives will be denoted by $\mathfrak{B}''$. In this case $\eta(x)$ is also of integrable square. In order to prove this, observe that the inequality

$$\begin{aligned} |\eta(x)|^2 &= \left| \int_{a_k}^{x_k} \dot\eta_k(t_k, x'_k)\, dt_k \right|^2 \\ &\leqq |x_k - a_k| \int_{a_k}^{b_k} |\dot\eta_k(t_k, x'_k)|^2 dt_k. \end{aligned}$$

holds almost everywhere on S provided a_k and b_k are chosen so that the projection of S on the x_k-axis is within the interval $[a_k, b_k]$. The right member is integrable on S and hence also $|\eta(x)|^2$, as was to be proved.

The third class of variations, which we denote by $\mathfrak{B}_0$, consists of all variations η in $\mathfrak{B}''$ satisfying the condition

$$(2.5) \qquad I_{\sigma 1}(\eta) = \int_S \{f_{\sigma y}\eta + f_{\sigma p_i}\dot\eta_i\}\, dx = 0 \qquad (\sigma = 1, \cdots, r),$$

the arguments in the derivatives of f_σ being $[x, y_0(x), \dot y_0(x)]$. The function $I_{\sigma 1}(\eta)$ is the *first variation* of $I_\sigma(y)$ on y_0. The class $\mathfrak{B}_0$ plays a dominant role in the study of the second variation of $I(y)$ on y_0.

It should be observed that because of property (ii) of a function $\eta(x)$ in $\mathfrak{B}'$ we have the relation

$$(2.6) \qquad \int_{a'_k}^{b'_k} \{\eta(b_k, x'_k) - \eta(a_k, x'_k)\}\, dx'_k = \int_a^b \dot\eta_k(x)\, dx \quad (k = 1, \cdots, m)$$

which holds on every interval $[a, b]$. Moreover the function $\xi(x) = \lim_{h=0} \eta^h(x)$, where η^h is the integral mean of $\eta(x)$, differs from $\eta(x)$ at most on a set of measure zero and belongs to the same class $\mathfrak{B}'$, $\mathfrak{B}''$ as η does. Consequently in our proof we can assume that $\eta(x)$ is equal to the limit of its integral mean. Functions of this latter type, that belong to $\mathfrak{B}'$ can be characterized as described in the following

LEMMA 2.1. *Let $\eta(x)$ be an integrable function on S that is identically zero on the complement of S and is equal to the limit of its integral mean whenever this limit exists. If there exist integrable functions $\dot\eta_k(x)$ $(k = 1, \cdots, m)$ such that the equation (2.6) holds for almost all intervals $[a, b]$, then $\eta(x)$ belongs to the class $\mathfrak{B}'$ and $\dot\eta_k(x)$ is equal almost everywhere to the derivative of $\eta(x)$ with respect to x_k.*

The proof of this lemma and of the remarks in the paragraph preceding this lemma can be found in the paper (4) by Calkin.

LEMMA 2.2. *Let $u_1(x), \cdots, u_m(x)$, $v(x)$ be continuous functions on S. The relation*

$$(2.7) \qquad \int_S (u_i\dot\eta_i + v\eta)\,dx = 0$$

holds for all functions η in $\mathfrak{B}'$ if and only if the formula

$$(2.8) \qquad \int_a^b v(x)\,dx = \int_{a'_i}^{b'_i} \{u_i(b_i, x'_i) - u_i(a_i, x'_i)\}\,dx'_i \qquad (i \text{ summed})$$

holds for every closed interval $[a, b]$ interior to S.

The necessity of the condition (2.8) has been established by Carson (5). He has also shown that the condition (2.8) implies (2.7) under more restrictive hypotheses than those here imposed. In order to prove that the condition (2.8) implies (2.7) as stated above, let the functions $u_i(x)$ be extended so as to be continuous for all values of x and let $u^h_i(x)$ be their integral means. Set $v^h(x) = \partial u^h_i/\partial x_i$ (i summed). Consider an interval $[A, B]$ containing S in its interior. By iterated integration we have

$$\int_A^B \{u^h_i\dot\eta_i + v^h\eta\}\,dx = \int_A^B (\partial/\partial x_i)(\eta u^h_i)\,dx = 0$$

for every variation η in $\mathfrak{B}'$, η being zero on the boundary of $[A, B]$. But $\eta(x) \equiv \dot\eta_k(x) \equiv 0$ exterior to S. It follows that

$$\int_S \{u^h{}_i\dot{\eta}_i + v^h\eta\}dx = 0. \tag{2.9}$$

By virtue of the relation (2.8) and the definition of $v^h(x)$ it is seen that, at each point interior to S, $v^h(x)$ is the integral mean of $v(x)$ provided h is sufficiently small. Consequently $\lim_{h=0} v^h(x) = v(x)$ interior to S. In view of this fact (2.7) follows from (2.9).

3. The sufficiency theorem. The principal theorem in the present paper is the sufficiency theorem described in this section. Its proof will be given later. In this theorem we impose certain conditions upon y_0 that will insure that y_0 will afford a proper strong relative minimum to $I(y)$ in the class $\mathfrak{A}_0$. These conditions are the customary necessary conditions suitably strengthened. They are as follows:

I. There exists a set of multipliers $\lambda_0 \geqq 0$, λ_σ, not all zero, such that given a closed interval $[a, b]$ in S the formula

$$\int_a^b F_y dx = \int_{a'_i}^{b'_i} F_{p_i} \Big|_{a_i}^{b_i} dx'_i \qquad (i \text{ summed}) \tag{3.1}$$

holds on y_0 with

$$F(x, y, p, \lambda) = \lambda_0 f(x, y, p) + \lambda_\sigma f_\sigma(x, y, p). \tag{3.2}$$

According to Lemma 2.2 this condition is equivalent to the condition that

$$J_1(\eta, \lambda) = \int_S \{F_{p_i}\dot{\eta}_i + F_y\eta) dx = 0 \tag{3.3}$$

for every variation η in $\mathfrak{B}'$. This condition is commonly called the *first necessary condition.*

II$_n$. The multipliers in I can be chosen so that there exists a neighborhood $\mathfrak{N}$ of y_0 in xyp-space and a constant $h > 0$ such that the inequality

$$E_F(x, y, p, q, \lambda) \geqq h \mid E_{f\sigma}(x, y, p, q) \mid \qquad (\sigma = 1, \cdots, r) \tag{3.4}$$

holds whenever (x, y, p) is in $\mathfrak{N}$ and (x, y, q) is in $\mathcal{R}$. Here E_F is the Weierstrass E-function

$$\begin{aligned} E_F(x, y, p, q, \lambda) &= F(x, y, q, \lambda) - F(x, y, p, \lambda) \\ &\quad - (q_i - p_i) F_{p_i}(x, y, p, \lambda) \end{aligned} \tag{3.5}$$

and $E_{f\sigma}$ is the corresponding E-function for $f_\sigma(x, y, p)$. This condition is called the *strengthened condition of Weierstrass.*

III′. The multipliers $\lambda_0 \geqq 0$, λ_σ in I and IIn can be chosen so that the inequality

$$(3.6) \qquad F_{p_i p_j}(x, y_0(x), \dot{y}_0(x), \lambda)\pi_i\pi_j > 0$$

holds for every set $\pi \neq 0$ and every x on S. If y_0 satisfies IIn this condition holds if and only if the matrix

$$(3.7) \qquad \| F_{p_i p_j}(x, y_0(x), \dot{y}_0(x), \lambda) \|$$

is nonsingular on the closure of S. This condition is called the *strengthened condition of Legendre.*

IV′. Given a non-null variation $\eta(x)$ in $\mathfrak{B}_0$ there exists a set of multipliers $\lambda_0 \geqq 0$, λ_σ such that I, IIn, III′ hold and such that the further inequality

$$(3.8) \qquad J_2(\eta, \lambda) = \int_S 2\omega(x, \eta, \dot{\eta}, \lambda)\, dx > 0,$$

holds, where

$$(3.9) \qquad 2\omega(x, \eta, \pi, \lambda) = F_{yy}\eta^2 + 2F_{yp_i}\eta\pi_i + F_{p_i p_j}\pi_i\pi_j,$$

the derivatives of $F(x, y, p, \lambda)$ being evaluated on y_0.

It is clear that y_0 cannot satisfy condition IV′ without satisfying the conditions I, IIn and III′.

The principal theorem to be established in the present paper is the following

THEOREM 3.1. *If the function y_0 satisfies the conditions* I, IIn, III′, IV′ *described above, there exists a neighborhood $\mathfrak{F}$ of y_0 in xy-space and a constant $\rho > 0$ such that the inequality*

$$(3.10) \qquad I(y) - I(y_0) \geqq \rho D(y - y_0)$$

holds for every function y in $\mathfrak{A}_0$ lying in $\mathfrak{F}$. Here $D(y)$ is defined by the formula

$$(3.11) \qquad D(y) = 2 \int_S \{V(\dot{y}) - 1\}\, dx$$

in which

$$(3.12) \qquad V(p) = (1 + p_i p_i)^{\frac{1}{2}}.$$

The function $D(y - y_0)$ is a measure of the deviation of y from y_0, and in fact equals twice the difference between the volume of the m-dimensional surface $\eta = y - y_0$ in xy-space and the surface $\eta \equiv 0$. The function $D(y)$ plays an important role in the proof of Theorem 3.1.

It may happen that there is a unique set of multipliers of the form $\lambda_0 = 1, \lambda_\sigma$ with which y_0 satisfies condition I. This case is called the *normal case.* All other cases are called *abnormal.* By the *completely abnormal case* will be meant the one in which the multipliers $\lambda_0 \geqq 0$, λ_σ in condition IV′ can be chosen so that $\lambda_0 = 0$ for each η in $\mathfrak{B}_0$. It is easily seen (cf. 8) that except possibly in the completely abnormal case one can always choose $\lambda_0 = 1$ in condition IV′ for each η in $\mathfrak{B}_0$.

THEOREM 3. 2. *Let y_0 satisfy conditions* I, IIn, III′, IV′. *In the completely abnormal case, there is a neighborhood $\mathfrak{F}$ of y_0 in xy-space such that y_0 is the only function in $\mathfrak{A}_0$ that lies in $\mathfrak{F}$.*

The proof of this result will be given at the end of the next section.

4. A related problem in the abnormal case. In the abnormal case our problem can be modified so as to obtain a more general result than that stated in Theorem 3. 1. This modification will be described in this section.

We assume that our problem is abnormal. Then there exists more than one set of multipliers satisfying condition I. Let $\lambda_{0\tau}, \lambda_{\sigma\tau}$ $(\tau = 1, \cdots, s)$ be a maximal set of linearly independent multipliers with $\lambda_{0\tau} = 0$ such that equation (3. 1) holds as stated for each of these sets and hence for any linear combination of these sets. Obviously, $s \leqq r$, the number of isoperimetric conditions (1. 2). By virtue of Lemma 2. 2 the equation

$$\lambda_{\sigma\tau} I_{\sigma 1}(\eta) = 0 \qquad (\tau = 1, \cdots, s) \tag{4.1}$$

holds for every variation η in $\mathfrak{B}'$. It follows that, if we select additional constants $u_{\sigma\theta}$ $(\theta = 1, \cdots, r - s)$ such that $| \lambda_{\sigma\tau} \; u_{\sigma\theta} | \neq 0$, then a variation η in $\mathfrak{B}'$ will satisfy the conditions $I_{\sigma 1}(\eta) = 0$ $(\sigma = 1, \cdots, r)$ if and only if it satisfies the conditions

$$J_{\theta 1}(\eta) = u_{\sigma\theta} I_{\sigma 1}(\eta) = 0 \qquad (\theta = 1, \cdots, r - s). \tag{4.2}$$

The class $\mathfrak{B}_0$ is accordingly completely determined by the $r - s$ conditions (4. 2). The function $J_{\theta 1}(\eta)$ is the first variation on y_0 of the integral

$$J_\theta(y) = \int_S g_\theta(x, y, \dot{y})\, dx = u_{\sigma\theta} I_\sigma(y) \tag{4.3}$$

in which we have set

$$g_\theta(x, y, p) = u_{\sigma\theta} f_\sigma(x, y, p). \tag{4.4}$$

From the remarks just made it is clear that as far as the variations η

in $\mathfrak{B}_0$ are concerned we can replace the original set of r isoperimetric conditions (1.2) by the smaller set

$$J_\theta(y) = J_\theta(y_0) \qquad (\theta = 1, \cdots, r - s).$$

In fact this replacement can be made even in the sufficiency theorem if we modify our concept of a minimum in the manner described in Theorem 4.1 below. The integral $I(y)$ must also be modified as follows: Recall first that in the completely abnormal case the multipliers in condition IV′ can be restricted to be of the form

$$\lambda_0 = 0, \qquad \lambda_{\sigma\tau} c_\tau \qquad (\tau = 1, \cdots, s), \tag{4.5}$$

where $\lambda_{\sigma\tau}$ are the multipliers described in the last paragraph and the c's are constants. If the problem is not completely abnormal there is one set of multipliers of the form $\lambda_0 = 1$, λ_σ effective as described in condition IV′ for some variation in $\mathfrak{B}_0$. Every set of multipliers having $\lambda_0 = 1$ is expressible in the form

$$\lambda_0 = 1, \qquad \lambda_\sigma + \lambda_{\sigma\tau} c_\tau \qquad (\tau = 1, \cdots, s) \tag{4.6}$$

and, as was remarked at the end of the last section, we can restrict the multipliers in condition IV′ to be of this type. Using the multipliers (4.5) or (4.6), according as our problem is completely abnormal or not, we define a new integral by the formula

$$J(y, c) = \int_S g(x, y, \dot{y}, c)\,dx = \lambda_0 I(y) + (\lambda_\sigma + \lambda_{\sigma\tau} c_\tau) I_\sigma(y), \tag{4.7}$$

in which it is agreed that $\lambda_\sigma = 0$ in the completely abnormal case. The integrand g is given by the formula

$$g(x, y, p, c) = \lambda_0 f(x, y, p) + (\lambda_\sigma + \lambda_{\sigma\tau} c_\tau) f_\sigma(x, y, p). \tag{4.8}$$

In terms of the new integral $J(y, c)$ the conditions I, IIn, III′, IV′ imposed upon y_0 take the following form:

I. Given a closed interval $[a, b]$ in S the formula

$$\int_a^b g_y\,dx = \int_{a'_i}^{b'_i} g_{p_i} \Big|_{a_i}^{b_i} dx_i \qquad (i \text{ summed}) \tag{4.9}$$

holds on y_0 for every element $c = (c_1, \cdots, c_s)$ or equivalently

$$J_1(\eta, c) = \int_S \{g_{p_i}\dot{\eta}_i + g_y \eta\}\,dx = 0 \tag{4.10}$$

holds for all η in $\mathfrak{B}'$ and for all c, the arguments in the derivatives of g being $(x, y_0(x), \dot{y}_0(x), c)$.

II$_{\text{N}}$. There exists an element $c = (c_1, \cdots, c_s)$ in c-space, a neighborhood $\mathfrak{N}$ of y_0 in xyp-space and a constant $h > 0$ such that the inequality

$$E_g(x, y, p, q, c) \geqq h \mid E_{f\sigma}(x, y, p, q) \mid \qquad (\sigma = 1, \cdots, r) \tag{4.11}$$

holds whenever (x, y, p) is in $\mathfrak{N}$ and (x, y, q) is in $\mathcal{R}$. Here E_g and $E_{f\sigma}$ are the E-functions for g and f_σ respectively. The set of all elements c of this type forms an open set C^*. In order to prove this we use the relation

$$E_g(x, y, p, q, c + k) = E_g(x, y, p, q, c) + \lambda_{\sigma\tau} k_\tau E_{f\sigma}(x, y, p, q).$$

Suppose that c is in C^* and choose $\mathfrak{N}$ and h such that (4.11) holds. Choose δ such that whenever $\mid k \mid < \delta$ one has $\mid \lambda_{\sigma\tau} k_\tau \mid < h/2r$. Then

$$\begin{aligned} E_g(x, y, p, q, c + k) &\geqq E_g(x, y, p, q) - h/2 \mid E_{f\sigma}(x, y, p, q) \mid \\ &\geqq h/2 \mid E_{f\sigma}(x, y, p, q) \mid. \end{aligned}$$

It follows that the δ-neighborhood of c is in C^*. Consequently C^* is open, as was to be proved.

III$'$. There exists an element c in C^* such that the inequality

$$g_{p_i p_j}(x, y_0(x), \dot{y}_0(x), c)\pi_i \pi_j > 0 \tag{4.12}$$

holds whenever $\pi \neq 0$ and x is on S. The set of elements c of this type forms an open subset C of C^*. The matrix

$$\| g_{p_i p_j}(x, y_0(x), \dot{y}_0(x), c) \| \tag{4.13}$$

is nonsingular whenever c is in C and x is on S.

IV$'$. Given a variation $\eta \neq 0$ in $\mathfrak{B}_0$ there is an element c in C such that the inequality

$$J_2(\eta, c) = \int_S 2\omega(x, \eta, \dot{\eta}, c)\, dx > 0 \tag{4.14}$$

holds, where

$$2\omega(x, \eta, \pi, c) = g_{yy}\eta^2 + 2g_{yp_i}\eta\pi_i + g_{p_i p_j}\pi_i\pi_j, \tag{4.15}$$

the arguments in the derivatives of g being $(x, y_0(x), \dot{y}_0(x), c)$. It will be seen in Theorem 9.1 below that the c's can be restricted to lie on a compact (bounded closed) subset C_0 of C.

The generalized sufficiency theorem for the abnormal case is given by the following

THEOREM 4.1. *If a function y_0 satisfies the conditions* I, IIn, III′, IV′, *there exists a bounded and closed subset C_0 of C, a neighborhood $\mathfrak{F}$ of y_0 in xy-space and a constant $\rho > 0$ such that given a function y in $\mathfrak{A}$ lying in $\mathfrak{F}$ and satisfying the isoperimetric conditions*

$$J_\theta(y) = J_\theta(y_0) \qquad (\theta = 1, \cdots, r - s), \tag{4.16}$$

there exists an element c in C_0 such that the inequality

$$J(y, c) - J(y_0, c) \geqq \rho D(y - y_0) \tag{4.17}$$

holds, where $D(\eta)$ is given by (3.11).

It should be noticed that the conclusion in the theorem is analogous to the condition on the second variation $J_2(\eta, c)$ in IV′. It will be seen in the course of the proof that any compact subset C_0 of C that is effective in condition IV′ as indicated in Theorem 9.1 below is effective in Theorem 4.1.

If a function $y(x)$ lies in $\mathfrak{F}$ and belongs to the class $\mathfrak{A}_0$ determined by the initial isoperimetric conditions (1.2), then it satisfies conditions (4.16) and $J(y) = \lambda_0 I(y)$ by virtue of (3.7). It follows that in the completely abnormal case $J(y, c) \equiv 0$ on $\mathfrak{A}_0$, since $\lambda_0 = 0$. The condition (4.17) then takes the form $D(y - y_0) = 0$, which can hold only in case $y = y_0$. Theorem 3.2 therefore holds and Theorem 3.1 is vacuously true. If the original problem is not completely abnormal, then $\lambda_0 = 1$ and $J(y, c) = I(y)$ on $\mathfrak{A}_0$. Consequently (3.10) follows from (4.17) and Theorem 3.1 holds as stated. Even in this case we have no guarantee that there is a function $y \neq y_0$ in $\mathfrak{A}_0$ that lies in $\mathfrak{F}$.

For the normal case Theorem 3.1 can be considered to be the special case of Theorem 4.1 in which there are no c's, provided we interpret the remarks concerning the c's to be vacuously true. This will be done in the sequel. Our main problem is then that of establishing Theorem 4.1.

5. A property of the Weierstrass E-function. In order to carry out the proof of Theorem 4.1 it will be convenient to make use of a property of the Weierstrass E-function which is a consequence of conditions IIn and III′ imposed upon y_0. This result is not new. It has been used by Reid (18) and others. It is a consequence of certain results obtained by the author (6) in considering the parametric problem of Bolza. For the sake of completeness we shall give a brief outline of a proof. This proof is novel in the sense that we make use of an identity on the E-function (see equation (5.2) below) that does not appear to have been used before in this connection. This identity has a simple geometric interpretation, which we shall omit.

THEOREM 5.1. *Let y_0 satisfy conditions* IIn, III′ *of* **4** *and let C_0 be a compact subset of the set C appearing in these conditions. There exists a neighborhood $\mathcal{R}_0$ of y_0 in xyp-space and a constant $h > 0$ such that the inequalities* (4.11) *and*

$$E_g(x, y, p, q, c) \geqq 2h\{V(q - p) - 1\} \tag{5.1}$$

hold, whenever c is in C_0, (x, y, p) is in $\mathcal{R}_0$ and (x, y, q) is in $\mathcal{R}$. The function $V(p)$ is given by (3.12) *and $g(x, y, p, c)$ by* (4.8). *The function E_g is the Weierstrass E-function for g.*

It is interesting to observe that the difference $V(q - p) - 1$ is the Weierstrass E-function for $V(q - p)$ with p held fast.

Observe first that there is a neighborhood $\mathcal{R}_1$ of y_0 in xyp-space and a constant $h > 0$ such that the inequality (5.1) holds whenever c is in C_0, and (x, y, p) and (x, y, q) are in $\mathcal{R}_1$. This follows from condition III′ by a continuity argument together with the customary application of Taylor's theorem to the Weierstrass E-function (cf. 6). Diminish $\mathcal{R}_1$ and h if necessary so that (4.11) holds when c is in C_0, (x, y, p) is in $\mathcal{R}_1$ and (x, y, q) is in $\mathcal{R}$. An application of the Heine-Borel theorem is effective here.

Let $\mathcal{R}_0$ be a second neighborhood of y_0 whose closure is in $\mathcal{R}_1$ and let ϵ be a positive constant such that if (x, y, p) is in $\mathcal{R}_0$ then $(x, y, p + \pi)$ is in $\mathcal{R}_1$ for every vector π with $|\pi| \leqq \epsilon$. Consider now a set (x, y, p, q, c) with (x, y, p) in $\mathcal{R}_0$, (x, y, q) in $\mathcal{R} - \mathcal{R}_1$, and c in C_0. Select a vector π and a positive constant k such that $q = p + k\pi$ and $|\pi| = \epsilon$. Obviously $k > 1$ since (x, y, q) is not in $\mathcal{R}_1$. Consider the identity

$$\begin{aligned} E_g(p, p + k\pi) = {} & E_g(p + \pi, p + k\pi) \\ & + kE_g(p, p + \pi) + (k - 1)E_g(p + \pi, p), \end{aligned} \tag{5.2}$$

in which only the significant variables are displayed. The first and last items are non-negative, by (4.11). It follows that

$$E_g(p, p + k\pi) \geqq kE_g(p, p + \pi) \geqq 2hk[V(\pi) - 1],$$

the last inequality being obtained from the fact that (5.1) holds with (x, y, p) in $\mathcal{R}_0$ and $(x, y, p + \pi)$ in $\mathcal{R}_1$. But for $k > 1$ and $|\pi| = \epsilon$ we have the further inequality

$$k[V(\pi) - 1] \geqq \epsilon'[V(k\pi) - 1], \qquad \epsilon' = \epsilon/(2 + \epsilon)$$

which follows readily from the definition (3.12) of $V(p)$. Consequently

$$E_g(p, q) \geqq 2h\epsilon'[V(q - p) - 1].$$

The formula (5.1) accordingly holds for a suitable choice of h, as was to be proved.

COROLLARY. *Let C_0 be a compact subset of C and let*

$$E^*(y,c)=\int_S E_g(x,y,\dot{y}_0(x),\dot{y},c)\,dx \tag{5.3}$$

$$E^*_\theta(y)=\int_S E_{g_\theta}(x,y,\dot{y}_0(x),\dot{y})\,dx \tag{5.4}$$

$$D(y)=2\int_S \{V(\dot{y})-1\}\,dx. \tag{5.5}$$

There exists a neighborhood $\mathfrak{F}$ of y_0 in xy-space and a constant $h>0$ such that the inequalities

$$E^*(y,c)\geqq hD(y-y_0) \tag{5.6}$$

$$E^*(y,c)\geqq h\,|\,E^*_\theta(y)\,| \qquad (\theta=1,\cdots,r-s) \tag{5.7}$$

hold for every element c in C_0 and every function y in $\mathfrak{A}$ lying in $\mathfrak{F}$.

This result follows at once from the inequalities (4.11) and (5.1) and the definition (4.4) of g_θ.

6. An analogue of a theorem of Lindeberg. The analogue of the theorem of Lindeberg given below is established under weaker hypotheses than those imposed on y_0 in the preceding sections. The principal hypotheses are as follows: Consider an integral

$$J(y,c)=\int_S g(x,y,\dot{y},c)\,dx,$$

where $g(x,y,p,c)$ is continuous and has continuous derivatives with respect to $y,p_1,\cdots,p_m$ for all (x,y,p) in $\mathfrak{R}$ and $c=(c_1,\cdots,c_s)$ in an open set C in c-space. It is assumed that given a compact subset C_0 of C there exists a neighborhood $\mathfrak{F}$ of y_0 in xy-space and positive constants α, β such that the first of the inequalities

$$E_g(x,y,\dot{y}_0(x),p,c)\geqq 0 \tag{6.1}$$

$$E_g(x,y,\dot{y}_0(x),p,c)\geqq \alpha\,|\,p-\dot{y}_0(x)\,| \tag{6.2}$$

holds whenever c is in C_0, (x,y) is in $\mathfrak{F}$, (x,y,p) is in $\mathfrak{R}$ and the second holds if in addition one has $|\,p-\dot{y}_0(x)\,|\geqq\beta$. The second condition can be replaced by an inequality of the form

$$E_g(x, y, \dot{y}_0(x), p, c) \geqq 2h\{V(p - \dot{y}_0(x)) - 1\} \qquad (h > 0)$$

subject to the restriction that $| p - \dot{y}_0(x) | \geqq \beta$. It follows that there is a constant N such that

$$D(y - y_0) \leqq N[1 + E^*(y, c)] \tag{6.3}$$

whenever c is in C_0 and y is a function in $\mathfrak{A}$ lying in $\mathfrak{F}$. The function $D(y)$ is defined by (5.5) and $E^*(y, c)$ is the corresponding integral (5.3) determined by the function $g(x, y, p, c)$ here used.

The proofs of the theorems in this section are based upon the following

LEMMA 6.1. *Let $I^*(y, c)$ be an integral of the form*

$$I^*(y, c) = \int_S \{u_i(x, y, c)\dot{y}_i + v(x, y, c)\}dx, \tag{6.4}$$

where $u_i(x, y, c)$, $v(x, y, c)$ are continuous functions of (x, y, c) for (x, y) on a neighborhood of those on y_0 and $c = (c_1, \cdots, c_s)$ on a compact set C_0. Given a constant $\epsilon > 0$ there exists a neighborhood $\mathfrak{F}$ of y_0 in xy-space such that the inequality

$$| I^*(y, c) - I^*(y_0, c) | < \epsilon[1 + D(y - y_0)] \tag{6.5}$$

holds whenever c is in C_0 and y is a function in $\mathfrak{A}$ lying in $\mathfrak{F}$. The function $D(y)$ is defined by equation (5.5).

In the proof we can assume without loss of generality that $y_0(x) \equiv 0$. From the definition of $D(y)$ it is clear that there is a constant Q such that

$$\int_S V(\dot{y})dx \leqq Q[1 + D(y)] \tag{6.6}$$

for every function y in $\mathfrak{A}$. Given a number $\epsilon > 0$ select functions $U_i(x, c)$ of class C' such that

$$| U(x, c) - u(x, 0, c) | < \epsilon/3Q$$

whenever x is in S and c is in C_0. Set

$$A_i(x, y, c) = u_i(x, y, c) - U_i(x, c)$$

$$B(x, y, c) = v(x, y, c) - v(x, 0, c) - y(\partial U_i/\partial x_i).$$

Choose a neighborhood $\mathfrak{F}$ of y_0 such that the inequality

$$A_iA_i + B^2 < \epsilon^2/Q^2 \tag{6.7}$$

holds whenever (x, y) is in $\mathfrak{F}$ and c is in C_0. Observe that $I^*(y, c)$ is expressible in the form

$$I^*(y, c) = J^*(y, c) + K^*(y, c)$$

where

$$J^*(y, c) = \int_S \{v(x, 0, c) + (\partial/\partial x_i)(yU_i(x, c))\}dx,$$

$$K^*(y, c) = \int_S \{A_i(x, y, c)\dot{y}_i + B(x, y, c)\}dx.$$

Consider now a function y in $\mathfrak{A}$ lying in $\mathfrak{F}$. Clearly $J^*(y, c) = J^*(y_0, c)$. Moreover, by (6.7) and the relation $K^*(y_0, c) = 0$ we have

$$|I^*(y, c) - I^*(y_0, c)| = |K^*(y, c)| \leqq (\epsilon/Q)\int_S V(\dot{y})dx.$$

Combining this result with (6.6) one obtains (6.5) with $y_0(x) \equiv 0$. This proves the lemma.

We have the following result concerning lower semicontinuity.

THEOREM 6.1. *Let C_0 be a compact subset of C and let ϵ be a positive constant. If the conditions (6.1) and (6.2) hold as stated, there exists a neighborhood $\mathfrak{F}$ of y_0 in xy-space such that the inequality*

$$J(y, c) \geqq J(y_0, c) - \epsilon \tag{6.8}$$

holds for every element c in C and every function y in $\mathfrak{A}$ lying in $\mathfrak{F}$.

In the proof of this result we use the familiar formula

$$J(y, c) = J^*(y, c) + E^*(y, c)$$

in which $E^*(y, c)$ is defined by the formula (4.3) and $J^*(y, c)$ is the Hilbert integral

$$J^*(y, c) = \int_S \{g(x, y(x), \dot{y}_0(x), c) + (\dot{y}_i(x) - \dot{y}_{0i}(x))g_{p_i}(x, y(x), \dot{y}_0(x), c)\}dx. \tag{6.9}$$

Since $J^*(y_0, c) = J(y_0, c)$ it follows that

$$J(y, c) - J(y_0, c) = J^*(y, c) - J^*(y_0, c) + E^*(y, c). \tag{6.10}$$

Let C_0 be a compact subset of C and let ϵ be a positive constant. We can suppose that $\epsilon < 1$. Choose $\mathfrak{F}$ so that the inequalities (6.1) and (6.2) hold

as stated and let N be the number appearing in (6.3). By virtue of Lemma 6.1 we can diminish $\mathfrak{F}$ so that the relation

$$|J^*(y,c) - J^*(y_0,c)| < (\epsilon/(N+1))[1 + D(y-y_0)]$$

holds, whenever c is in C_0 and y is a function in $\mathfrak{A}$ lying in $\mathfrak{F}$. Using (6.3) we find that

$$|J^*(y,c) - J^*(y_0,c)| < \epsilon[1 + E^*(y,c)].$$

Consequently (6.10) yields

$$J(y,c) - J(y_0,c) \geqq -\epsilon + (1-\epsilon)E^*(y,c).$$

The last term is nonnegative by (6.1). It follows that the relation (6.8) holds, as was to be proved.

Consider now a second integral

$$K(y,c) = \int_S G(x,y,\dot{y},c)\,dx$$

of the same type as $J(y,c)$. We assume that given a compact subset C_0 of C one can select a neighborhood $\mathfrak{F}$ of y_0 in xy-space and a constant $h > 0$ such that the inequality

$$E_g(x,y,\dot{y}_0(x),p,c) \geqq h\,|E_G(x,y,\dot{y}_0(x),p,c)| \tag{6.11}$$

holds whenever c is in C_0, (x,y) is in $\mathfrak{F}$ and (x,y,p) is in $\mathfrak{R}$. Again E_g, E_G are the E-functions for g, G respectively.

The following theorem can be considered as an analogue of the theorem of Lindeberg.

THEOREM 6.2. *Suppose that the conditions* (6.1), (6.2), (6.11) *hold as stated and let C_0 be a compact subset of C. There exists a constant $\rho > 0$ such that given a number $\epsilon > 0$ there is a neighborhood $\mathfrak{F}$ of y_0 in xy-space for which the inequality*

$$J(y,c) - J(y_0,c) \geqq \rho[\,|K(y,c) - K(y_0,c)| - \epsilon] \tag{6.12}$$

holds, whenever c is in C_0 and y is a function in $\mathfrak{A}$ lying in $\mathfrak{F}$.

In order to establish this result consider the integral

$$I(y,c,t) = \int_S F(x,y,\dot{y},c,t)\,dx = J(y,c) + tK(y,c) \tag{6.13}$$

in which

$$F(x,y,p,c,t) = g(x,y,p,c) + tG(x,y,p,c).$$

Given a compact subset C_0 of C choose $\mathfrak{F}$, h, α, β so that conditions (6.1), (6.2), (6.11) hold as stated. Set $\rho = h/2$. A simple argument will show that for $t = \pm\rho$ the conditions (6.1) and (6.2) will remain valid if g is replaced by F and α by $\alpha/2$. Consequently we may apply Theorem 6.1 to $I(y, c, t)$ using $\epsilon' = \epsilon\rho$ in place of ϵ. It follows that we can diminish $\mathfrak{F}$ so that the inequality

$$I(y, c, \pm\rho) - I(y_0, c, \pm\rho) \geqq -\epsilon\rho$$

will hold whenever c is in C_0 and y is a function in $\mathfrak{A}$ lying in $\mathfrak{F}$. This inequality together with (6.13) yields the desired relation (6.12).

Observe that the E-function for the integral $D(y - y_0)$ is given by $V(p - \dot{y}_0(x)) - 1$. Consequently under the hypotheses of **4** it follows from Theorem 5.1 that the inequality (6.11) holds when $K(y, c) = D(y - y_0)$. For this special case Theorem 6.2 takes the special form described in the following

THEOREM 6.3. *Let $J(y, c)$ be the function described in* **4** *and suppose that y_0 satisfies conditions* IIn, III′. *Let C_0 be a compact subset of C. There exists a number $\rho > 0$ such that given a number $\epsilon > 0$ there is a neighborhood $\mathfrak{F}$ of y_0 in xy-space such that the inequality*

$$J(y, c) - J(y_0, c) \geqq \rho[D(y - y_0) - \epsilon] \tag{6.14}$$

holds whenever c is in C_0 and y is a function in $\mathfrak{A}$ lying in $\mathfrak{F}$.

7. Auxiliary lemmas. The present section will be devoted to a set of lemmas which will be useful in the proofs of the convergence theorems given in the next section. We begin with the following

LEMMA 7.1. *If η is a function in the class $\mathfrak{B}'$ described in* **2**, *satisfying a condition of the form*

$$\int_S |\dot{\eta}(x)|\, dx < N, \tag{7.1}$$

then its integral mean η^h satisfies the inequality

$$\int_S |\eta^h(x) - \eta(x)|\, dx < mNh. \tag{7.2}$$

In order to prove this inequality observe that

$$\begin{aligned} \eta^h(x) - \eta(x) &= (2h)^{-m} \int_{-h}^{h} \{\eta(x+t) - \eta(x)\} dt \\ &= (2h)^{-m} \int_{-h}^{h} \int_0^1 t_i \dot{\eta}_i(x + \theta t)\, d\theta dt. \end{aligned} \tag{7.3}$$

Hence

$$(7.4) \qquad \int_S |\eta^h(x) - \eta(x)|\, dx$$
$$\leqq (2h)^{-m} \int_0^1 \int_{-h}^{h} |t_i| \int_S |\dot\eta_i(x+\theta t)|\, dx\, dt\, d\theta.$$

Since $\eta \equiv \dot\eta_i \equiv 0$ exterior to S we have, by (7.1)

$$\int_S |\dot\eta_i(x+\theta t)|\, dx < N.$$

Using this inequality in (7.4) one obtains (7.2) by integration.

LEMMA 7.2. *If η is a function in the class $\mathfrak{B}''$ described in* **2** *which satisfies a condition of the form*

$$(7.5) \qquad \int_S |\dot\eta(x)|^2 dx < N,$$

then its integral mean η^h satisfies the inequality

$$(7.6) \qquad \int_S |\eta^h(x) - \eta(x)|^2 dx < mNh^2.$$

For by (7.3) and the inequality of Schwarz it is seen that

$$|\eta^h(x) - \eta(x)|^2 \leqq (2h)^{-2m} \Big[\int_{-h}^{h} \int_0^1 |\dot\eta_i(x+\theta t)|^2 dt\Big]\Big[\int_{-h}^{h} |t_i|^2 dt\Big]$$

Evaluating the last integral it is seen that

$$\int_S |\eta^h(x) - \ (x)|^2 dx \leqq m2^{-m}h^{-m+2} \int_{-h}^{h} \int_0^1 \int_S |\dot\eta(x+\theta t)|^2 dx\, d\theta\, dt.$$

Using the relation (7.5) together with the fact that $\dot\eta_k \equiv 0$ exterior to S it is found that (7.6) holds as was to be proved.

The next lemma has been established by McShane and Reid (18).

LEMMA 7.3. *If $\phi(t)$ is absolutely continuous on $a \leqq t \leqq b$ and $|\phi(t)| < \delta$, then*

$$(7.7) \qquad \int_a^b |\phi\phi'|\, dt \leqq d_1\Big[\int_a^b R(\phi')\, dt + \phi(a)^2\Big]$$

$$(7.8) \qquad \int_a^b \phi^2 dt \leqq d_2\Big[\int_a^b R(\phi')\, dt + \phi(a)^2\Big]$$

where d_1, d_2 are constants depending on δ and $b - a$ and $R(z) = \sqrt{1+z^2} - 1$.

We shall use this lemma to establish the further result

LEMMA 7.4. *If $\eta(x)$ is a function in $\mathfrak{B}'$ having $|\eta(x)| < \delta$ on S, then*

$$\int_S |\eta| \; |\dot{\eta}| \, dx \leqq \delta_1 D(\eta) \tag{7.9}$$

$$\int_S \eta^2 dx \leqq \delta_2 D(\eta) \tag{7.10}$$

where $D(\eta)$ is given by (3.11) and δ_1, δ_2 are constants depending upon δ and S.

In the proof we can assume that S is an interval $[a, b]$. Then $\eta(x_k, x'_k)$ is absolutely continuous in x_k on $[a_k, b_k]$ for almost all x'_k on $[a'_k, b'_k]$. Moreover $\eta(a_k, x'_k) = 0$. Using Lemma 7.3 with $t = x_k$ it is seen that the relation

$$\int_{a_k}^{b_k} | \eta(x_k, x'_k) | \; | \dot{\eta}_k(x_k, x'_k) | \, dx_k \leqq d_1 \int_{a_k}^{b_k} R[\dot{\eta}_k(x_k, x'_k)] dx_k$$

holds for almost all x'_k on $[a'_k, b'_k]$. Consequently

$$\int_a^b | \eta | \; | \dot{\eta}_k | \, dx \leqq d_1 \int_a^b R(\dot{\eta}_k) dx \leqq d_1 \int_a^b R(| \dot{\eta} |) dx = d_1 D(\eta).$$

The relation (7.9) therefore holds with $\delta_1 = md_1$. The inequality (7.10) follows in a similar manner from (7.8).

We conclude this section with two lemmas on weak convergence. A sequence of functions $\{z_q(x)\}$ is said to *converge weakly* to a function $z_0(x)$ in the class $\mathfrak{L}_\alpha$ $(\alpha \geqq 1)$ of Lebesgue integrable functions on S if $z_0, z_1 \cdots$ are in $\mathfrak{L}_\alpha$ and

$$\lim_{q=\infty} \int_S A(x) z_q(x) dx = \int_S A(x) z_0(x) dx \tag{7.11}$$

for every function $A(x)$ in the Lebesgue class $\mathfrak{L}_\beta$ where $\beta = (\alpha - 1)/\alpha$ if $\alpha > 1$ and $\beta = \infty$ if $\alpha = 1$.

LEMMA 7.5. *Let $\{z_q(x)\}$ be a sequence of functions in $\mathfrak{L}_\alpha$ $(\alpha \geqq 1)$ satisfying a condition of the form*

$$\int_S | z_q(x) |^\alpha dx < N. \tag{7.12}$$

In case $\alpha = 1$ assume further that the sequence of set functions

$$\int_M | z_q(x) | \, dx$$

is uniformly absolutely continuous on S. There exist a subsequence, again denoted by $\{z_q(x)\}$, and a function z_0 in $\mathfrak{L}_a$ such that z_q converges weakly to z_0 in $\mathfrak{L}_1$ on S.

As a second lemma we have

LEMMA 7.6. *If a sequence of functions $\{z_q(x)\}$ converges weakly to a function $z_0(x)$ in $\mathfrak{L} = \mathfrak{L}_1$ on S and the inequality (7.12) holds, then the sequence converges weakly to $z_0(x)$ in $\mathfrak{L}_1$ on S.*

For a proof of these Lemmas on weak convergence for the case $m = 1$ see the treatise by Banach (1), pp. 126-132.

8. Convergence theorems. We are now in position to prove convergence theorems that will be useful in the proof of Theorem 4.1 and in the study of the second variation. The classes $\mathfrak{B}'$ and $\mathfrak{B}''$ found in these theorems are defined in **2**.

THEOREM 8.1. *Let $\xi_0, \xi_1, \xi_2, \cdots$ be a sequence of functions in $\mathfrak{B}'$ satisfying the relations*

$$\lim_{q=\infty} \int_S |\xi_q(x) - \xi_0(x)|\, dx = 0, \qquad \int_S |\xi_q|\, dx < N \tag{8.1}$$

where N is a constant. If $P(x), Q_1(x), \cdots, Q_m(x)$ are continuous functions on S then

$$\lim_{q=\infty} \int_S \{P\xi_q + Q_i\dot{\xi}_{qi}\} dx = \int_S \{P\xi_0 + Q_i\dot{\xi}_{0i}\} dx. \tag{8.2}$$

Observe first that the equation (8.1) implies that

$$\lim_{q=\infty} \int_S P(\xi_q - \xi_0)\, dx = 0 \tag{8.3}$$

for every bounded integrable function P. If $Q(x)$ is a function of class C' for all values of x, then given an interval $[a, b]$ containing S we have

$$\int_S Q(\dot{\xi}_{qk} - \dot{\xi}_{0k})\, dx = \int_a^b Q(\dot{\xi}_{qk} - \dot{\xi}_{0k})\, dx = -\int_a^b (\partial Q/\partial x_k)(\xi_q - \xi_0)\, dx$$
$$= -\int_S (\partial Q/\partial x_k)(\xi_q - \xi_0)\, dx.$$

In this formula we have used freely the relations $\xi_q \equiv \xi_0 \equiv 0$ exterior to S. It follows from (8.3) that

$$\lim_{q=\infty} \int_S Q(\dot{\xi}_{qk} - \dot{\xi}_{0k})\, dx = 0$$

in this case. Consider now a function $P(x)$ that is continuous on S. Let ϵ be a positive constant. Select a function $Q(x)$ of class C' such that $| P(x) - Q(x) | < \epsilon/4$ on S. Then by virtue of the inequality in (8.1) we have

$$| \int_S (P - Q)(\dot{\xi}_{qk} - \dot{\xi}_{0k})\, dx | \leqq (\epsilon/4N) [\int_S | \dot{\xi}_q |\, dx + \int_S | \dot{\xi}_0 |\, dx] < \epsilon/2.$$

Next choose a number q_0 such that when $q \geqq q_0$ we have

$$| \int_S Q(x)(\dot{\xi}_{qk} - \dot{\xi}_{0k}) | < \epsilon/2.$$

Using the relation

$$\int_S P(\dot{\xi}_{qk} - \dot{\xi}_{0k})\, dx = \int_S (P - Q)(\dot{\xi}_{qk} - \dot{\xi}_{0k}) + \int_S Q(\dot{\xi}_{qk} - \dot{\xi}_{0k})\, dx$$

it follows that

$$| \int_S P(\dot{\xi}_{qk} - \dot{\xi}_{0k})\, dx | < \epsilon.$$

Hence

$$\lim_{q=\infty} \int_S P(\dot{\xi}_{qk} - \dot{\xi}_{0k})\, dx = 0$$

in this case also. This proves Theorem 8.1.

THEOREM 8.2. *Let $\{\xi_q\}$ be a sequence of functions in $\mathfrak{B}'$ satisfying a condition of the form*

$$(8.4) \qquad \int_S | \dot{\xi}_q(x) |\, dx < N \qquad (q = 1, 2, \cdots).$$

There exists a subsequence $\{\eta_q\}$ that converges almost everywhere on S to a function η_0 in $\mathfrak{L}$ and is such that

$$(8.5) \qquad \lim_{q=\infty} \int_S | \eta_q - \eta_0 |\, dx = 0.$$

In order to prove this result we can assume that S is an interval $[a, b]$, since $\xi_q \equiv 0$ exterior to S. Then on $[a_k, b_k]$ we have

$$| \xi_q(x_k, x'_k) | = | \int_{a_k}^{x_k} \dot{\xi}_{qk}(t_k, x'_k)\, dt_k | \leqq \int_{a_k}^{b_k} | \dot{\xi}_q |\, dt_k$$

for almost all x'_k on $[a'_k, b'_k]$. Consequently

$$(8.6) \qquad \int_a^b |\xi_q|\,dx \leqq (b_k - a_k) \int_a^b |\dot{\xi}_q|\,dx \leqq (b_k - a_k)N.$$

Consider now the integral mean ξ_q^h of ξ_q. By virtue of (8.4) we have

$$|\dot{\xi}_{qk}^h| = |(2h)^{-m} \int_{-h}^{h} \dot{\xi}_{qk}(x+t)\,dx| \leqq N(2h)^{-m}.$$

Similarly from (8.6) we have

$$|\xi_q^h| \leqq (b_k - a_k)N(2h)^{-m}.$$

It follows from Ascoli's Theorem that the sequence $\{\xi_q^h\}$ contains a uniformly convergent subsequence $\{\eta_q^h\}$. Let $\{\eta_q\}$ be the corresponding subsequence of $\{\xi_q\}$. Restrict h to be of the form $h = 2^{-p}$, where p is a positive integer. Then by means of a diagonal process a sequence $\{\eta_q\}$ can be chosen to be the same for each h on this restricted range.

Consider now a number $\epsilon > 0$. Choose $h = 2^{-p}$ such that $h < \epsilon/3mN$. Then by Lemma 7.1 we have

$$(8.7) \qquad \|\eta_q^h - \eta_q\| < \epsilon/3,$$

where for the moment

$$\|\eta\| = \int_S |\eta|\,dx.$$

Since the sequence η_q^h converges uniformly on S there is an integer q_0 such that whenever $q > r > q_0$ we have

$$(8.8) \qquad \|\eta_q^h - \eta_r^h\| < \epsilon/3.$$

Using the relation

$$(8.9) \qquad \|\eta_q - \eta_r\| \leqq \|\eta_q - \eta_q^h\| + \|\eta_q^h - \eta_r^h\| + \|\eta_r^h - \eta_r\|$$

it is seen that $\|\eta_q - \eta_r\| < \epsilon$ whenever $q > r > q_0$. Since the space $\mathfrak{L}$ of Lebesgue integrable functions is complete with respect to the norm $\|\eta\|$ it follows that there is a function η_0 in $\mathfrak{L}$ such that (8.5) holds.

THEOREM 8.3. *Let $\{\xi_q\}$ be a sequence of functions in $\mathfrak{B}'$ satisfying the condition* (8.4) *for a suitable choice of N. Suppose further that the sequence of set functions*

$$\int_M |\dot{\xi}_q|\,dx \qquad (q = 1, 2, \cdots)$$

is uniformly absolutely continuous on S. Then there exists a function η_0 in $\mathfrak{B}'$ such that (8.5) *holds. Moreover the subsequence can be selected so that η_q, $\dot\eta_{qk}$ $(k=1,\cdots,m)$ converge weakly in $\mathfrak{L}$ to η_0, $\dot\eta_{0k}$ respectively.*

In order to prove this result select $\{\eta_q\}$ and η_0 to have the properties described in Theorem 8.2. Then (8.5) holds and η_q converges weakly in $\mathfrak{L}$ to η_0. We can suppose that η_0 is equal to the limit of its integral mean when this limit exists, and that $\eta_0 = 0$ elsewhere. We shall show presently that η_0 is in $\mathfrak{B}'$. As a first step observe that the function $\dot\eta_{qk}$ satisfies the hypotheses of Lemma 7.5 with $\alpha = 1$. Consequently we can replace $\{\eta_q\}$ by a suitable subsequence such that $\dot\eta_{qk}$ converges weakly in $\mathfrak{L}$ to a function $\dot\eta_{0k}$. Consider now the equations

$$(8.10) \qquad \int_{a'_k}^{b'_k}\{\eta_q(b_k, x'_k) - \eta_q(a_k, x'_k)\}dx'_k = \int_a^b \dot\eta_{qk}dx$$

which hold for every interval $[a, b]$. Since $\dot\eta_{qk}$ converges weakly to $\dot\eta_{0k}$ the right member of this equation converges to the right member of the equation

$$(8.11) \qquad \int_{a'_k}^{b'_k}\{\eta_0(b_k, x'_k) - \eta_0(a_k, x'_k)\}dx'_k = \int_a^b \dot\eta_{0k}dx.$$

Since (8.5) holds, then except for a_k and b_k on a set M_k of measure zero in x_k-space the left member in (8.10) will converge in subsequence to the left member of (8.11). It follows that (8.11) holds unless a_k, b_k are on M_k. By virtue of Lemma 2.1 the function η_0 is in $\mathfrak{B}'$, as was to be proved.

THEOREM 8.4. *Let $\{\xi_q\}$ be a sequence of functions in $\mathfrak{B}''$ satisfying the condition*

$$(8.13) \qquad \int_S |\dot\xi_q|^2 dx < N.$$

There exists a subsequence $\{\eta_q\}$ and a function η_0 in $\mathfrak{B}''$ such that η_q, $\dot\eta_{qk}$ converge weakly on $\mathfrak{L}_2$ to η_0, $\dot\eta_{0k}$ respectively and

$$(8.14) \qquad \lim_{q=\infty} \int_S |\eta_q - \eta_0|^2 dx = 0.$$

In order to prove this result let M be a measurable subset of S. By the inequality of Schwarz we have

$$\int_M |\dot\xi_q|\, dx \leqq [\int_M |\dot\xi_q|^2 dx \int_M dx]^{\frac{1}{2}} \leqq |M|^{\frac{1}{2}} N^{\frac{1}{2}}$$

where $|M|$ is the measure of M. The hypotheses of Theorem 8.3 are satisfied and we can select a subsequence $\{\eta_q\}$ and a function η_0 in $\mathfrak{B}'$ having the properties described in that theorem. Since $\dot{\eta}_{qk}$ converges weakly in $\mathfrak{L}$ to $\dot{\eta}_{0k}$ and (8.13) holds with $\xi_q = \eta_q$ it follows that $\dot{\eta}_{qk}$ converges weakly to $\dot{\eta}_{0k}$ in $\mathfrak{L}_2$. Consequently η_{0k} is in $\mathfrak{B}''$. That η_q converges weakly to η_0 in $\mathfrak{L}_2$ will follow from the relation (8.14). To prove (8.14) we may suppose as in the proof of Theorem 8.2 that the sequence $\{\eta_q\}$ has been selected so that its integral mean $\eta_q{}^h$ converges uniformly on S at least if we restrict h to be of the form $h = 2^{-p}$, where p is an integer. By Lemma 7.2 we can select h so small that (8.7) holds with

$$\| \eta \| = \int_S | \eta_q |^2 dx$$

for a preassigned number $\epsilon > 0$. Next choose q_0 such that (8.8) holds with $q > r > q_0$. By (8.9) we then have $\| \eta_q - \eta_r \| < \epsilon$ if $q > r > q_0$. Since $\lim_{q=\infty} \eta_q = \eta_0$ almost everywhere it follows that (8.14) holds, as was to be proved.

9. Properties of the second variation. The second variation of the problem described in **3** is of the form

$$J_2(\eta, c) = \int_S 2\omega(x, \eta, \dot{\eta}, c)\, dx,$$

where

$$2\omega(x, \eta, \pi, c) = P(x, c)\eta^2 + 2Q_i(x, c)\eta\pi_i + R_{ik}(x, c)\pi_i\pi_k.$$

The functions $P(x, c)$, $Q_i(x, c)$, $R_{ik}(x, c) = R_{ki}(x, c)$ are continuous in (x, c) for all points $x = (x_1, \cdots, x_m)$ and $c = (c_1, \cdots, c_s)$. Moreover there is an open set C such that the inequality

$$(9.1) \qquad R_{ij}(x, c)\pi_i\pi_j > 0$$

holds whenever $\pi \neq 0$, c is in C and x is on S. If c is restricted to lie on a compact subset C_0 of C one can select a constant h such that

$$(9.2) \qquad R_{ij}(x, c)\pi_i\pi_k \geqq h\pi_i\pi_j$$

for all $\pi \neq 0$ and all x on a neighborhood of S.

We begin with the following lemma

LEMMA 9.1. *Let $\{\eta_q\}$ be a sequence of variations in $\mathfrak{B}''$ satisfying the*

hypotheses and the conclusions of Theorem 8.4 *and let* η_0 *be its limit function. Then for each element* c *in* C *one has*

$$\liminf_{q=\infty} J_2(\eta_q, c) \geqq J_2(\eta_0, c). \tag{9.3}$$

Moreover the conditions

$$\eta_0(x) \equiv 0 \text{ on } S, \qquad \liminf_{q=\infty} J_2(\eta_q, c) \leqq 0 \tag{9.4}$$

can hold only in case

$$\liminf_{q=\infty} \int_S |\dot{\eta}_q|^2 dx = 0. \tag{9.5}$$

In view of Theorem 8.1 with $\xi_q = \eta_q^2$ the inequality (9.3) will be established if we show that

$$\liminf_{q=\infty} \int_S R_{ij}\dot{\eta}_{qi}\dot{\eta}_{qj} dx \geqq \int_S R_{ij}\dot{\eta}_{0i}\dot{\eta}_{0j}. \tag{9.6}$$

To this end we use the formula

$$R_{ij}\dot{\eta}_{qi}\dot{\eta}_{qj} = R_{ij}\dot{\eta}_{0i}\dot{\eta}_{0j} + 2R_{ij}\dot{\eta}_{0i}(\dot{\eta}_{qj} - \dot{\eta}_{0j}) + R_{ij}(\dot{\eta}_{qi} - \dot{\eta}_{0i})(\dot{\eta}_{qj} - \dot{\eta}_{0j}).$$

By virtue of (9.1) we have

$$\int_S R_{ij}\dot{\eta}_{qi}\dot{\eta}_{qj} dx \geqq \int_S R_{ij}\dot{\eta}_{0i}\dot{\eta}_{0j} dx + 2\int_S R_{ij}\dot{\eta}_{0i}(\dot{\eta}_{qj} - \dot{\eta}_{0j})\, dx.$$

Since $\dot{\eta}_{qj}$ converges weakly to $\dot{\eta}_{0j}$ in $\mathfrak{L}_2$ the last term has zero as its limit. It follows that (9.6) holds, as desired.

Suppose next that (9.4) holds. Then by Theorem 8.1 with $\eta_0 \equiv 0$ and equation (9.2) we have

$$0 \geqq \liminf_{q=\infty} J_2(\eta_q, c) = \liminf_{q=\infty} \int_S R_{ij}\dot{\eta}_{qi}\dot{\eta}_{qj} dx \geqq h \liminf_{q=\infty} \int_S |\dot{\eta}_q|^2 dx$$

It follows that (9.5) holds. This completes the proof of Lemma 9.1.

THEOREM 9.1. *Suppose that for every non null function* η *in* $\mathfrak{B}_0$, *there is an element* c *in* C *such that* $J_2(\eta, c) > 0$. *Then there exists a compact subset* C_0 *of* C *such that given a non null function* η *in* $\mathfrak{B}_0$ *there is an element* c *in* C_0 *such that* $J_2(\eta, c) > 0$.

Let $\{C_q\}$ be a sequence of compact subsets of C whose union is C and which has $C_q < C_{q+1}$. If the conclusion in Theorem 9.1 were false, there would exist for each integer q a variation $\eta_q \neq 0$ in $\mathfrak{B}_0$ such that $J_2(\eta_q, c) \leqq 0$ whenever c is in C_q. We can suppose that η_q has been normalized so that

$$\int_S |\dot{\eta}_q(x)|^2 dx = 1. \tag{9.7}$$

According to Theorem 8.4 we can replace $\{\eta_q\}$ by a subsequence, again denoted by $\{\eta_q\}$, having the properties described in Theorem 8.4. Let η_0 be its limit function. Since η_q is in $\mathfrak{B}_0$ we have $I_{\sigma 1}(\eta_q) = 0$, where $I_{\sigma 1}$ is given by (2.5). By Theorem 8.1 we have $I_{\sigma 1}(\eta_0) = 0$. It follows that η_0 is in $\mathfrak{B}_0$. Consider now an element c in C. Choose p such that c is in C_p. For $q \geqq p$ we have $J_2(\eta_q, c) \leqq 0$ and hence $J_2(\eta_0, c) \leqq 0$, by (9.3). If $\eta_0 \equiv 0$ the condition (9.4) would be satisfied and (9.5) would hold. This is impossible in view of (9.7). It follows that Theorem 9.1 holds as stated.

THEOREM 9.2. *Suppose that the hypotheses described in Theorem* 9.1 *hold and choose* C_0 *as in Theorem* 9.2. *Let* $K_2(\eta, c)$ *be a second function of the same form as* $J_2(\eta, c)$. *There exists a constant* $\rho > 0$ *such that given a variation* $\eta \neq 0$ *in* $\mathfrak{B}_0$ *there exists an element* c *in* C_0 *such that*

$$J_2(\eta, c) \geqq \rho \, |K_2(\eta, c)|. \tag{9.8}$$

Observe first that since $K_2(\eta, c)$ is of the same form as $J_2(\eta, c)$ we can select a constant N such that

$$|K_2(\eta, c)| \leqq N \int_S \{|\eta|^2 + |\dot{\eta}|^2\} dx \tag{9.9}$$

whenever c is in C_0. Suppose now that the theorem is false. Then there exists for every integer q a variation $\eta_q \neq 0$ in $\mathfrak{B}_0$ satisfying the inequality

$$J_2(\eta_q, c) \leqq (1/q) \, |K_2(\eta_q, c)|$$

whenever c is in C_0. We can suppose that η_q has been chosen so that

$$\int_S \{\eta^2_q + |\dot{\eta}_q|^2\} dx = 1 \tag{9.10}$$

and such that the sequence $\{\eta_q\}$ converges to a variation η_0 in $\mathfrak{B}_0$ in the sense described in Theorem 8.4. Then by (9.9) we have

$$0 \geqq \liminf_{q=\infty} J_2(\eta_q, c) \geqq J_2(\eta_0, c)$$

whenever c is in C_0. Consequently $\eta_0 = 0$ by Theorem 9.1. The relations (9.4) are accordingly satisfied. It follows that (9.5) holds. Using (8.14) with $\eta_0 \equiv 0$ it is seen from (9.10) that

$$\lim_{q=\infty} \int_S |\dot{\eta}_q|^2 dx = 1,$$

contrary to (9.5). The conclusion in Theorem 9.2 is accordingly valid, as was to be proved.

10. Proof of Theorem 4.1. Theorem 4.1 will be established by showing that one of its contrapositives is true. The particular one that we shall use is the following

THEOREM 10.1. *Suppose that y_0 satisfies the conditions* I, IIn *and* III′ *described in* **4** *and let C_0 be a compact subset of the set C appearing in these conditions. Suppose further that for every number $\rho > 0$ and every neighborhood $\mathfrak{F}$ of y_0 in xy-space there is a function y in $\mathfrak{A}$ lying in $\mathfrak{F}$ satisfying the conditions*

$$J(y,c) - J(y_0,c) < \rho D(y - y_0)$$

$$J_\theta(y) = J_\theta(y_0) \qquad (\theta = 1, \cdots, r-s)$$

for every c in C_0. Then there exists a variation $\eta_0 \neq 0$ in $\mathfrak{B}_0$ such that $J_2(\eta_0, c) \leqq 0$ whenever c is in C_0. Here, $J_2(\eta, c)$ is the second variation of $J(y,c)$ on y_0.

If we assume that this result is true, we can establish Theorem 4.1 as follows. Select C_0 as described in Theorem 9.1. Let η be a variation in $\mathfrak{B}_0$. By virtue of Theorem 9.1 the relation $J_2(\eta, c) \leqq 0$ holds for every element c in C_0 if and only if $\eta = 0$. It follows that the hypotheses of Theorem 10.1 and 4.1 are incompatible. Consequently Theorem 4.1 is true as stated.

The remainder of this section will be devoted to the proof of Theorem 10.1. By virtue of our hypotheses there exists a sequence $\{y_q(x)\}$ of functions in $\mathfrak{A}$ converging uniformly to $y_0(x)$ and satisfying the condition

$$(10.1) \qquad J(y_q,c) - J(y_0,c) < (1/q)D(y_q - y_0), \qquad J_\theta(y_q) = J_\theta(y_0)$$

whenever c is in C_0. From the first relation it follows that $y_q \neq y_0$. By virtue of Theorem 6.3 there is a constant $\rho > 0$ such that whenever c is in C_0 we have

$$\liminf\,[J(y_q,c) - J(y_0,c)] \geqq \rho[\liminf_{q=\infty} D(y_q - y_0) - \epsilon]$$

for every number $\epsilon > 0$ and hence also for $\epsilon = 0$. The last two inequalities can hold only in case

$$(10.2) \qquad \lim_{q=\infty} J(y_q,c) = J(y_0,c) \qquad (c \text{ in } C_0)$$

$$(10.3) \qquad \lim_{q=\infty} D(y_q - y_0) = 0.$$

From the definition of $D(y)$ it follows that the sequence $\{\dot{y}_{qk}\}$ converges to $\dot{y}_{0k}$ in measure. Replacing our original sequence by a suitable subsequence it can be brought about that $\lim_{q=\infty} \dot{y}_{qk}(x) = \dot{y}_{0k}(x)$ $(k = 1, \cdots, m)$ almost uniformly on S. Setting

$$\delta y_q = y_q - y_0, \qquad v_q = (\tfrac{1}{2})[1 + (1 + |\delta\dot{y}_q|^2)^{\frac{1}{2}}]$$

it is seen that $\lim_{q=\infty} v_q(x) = 1$ almost uniformly on S and that

$$D(\delta y_q) = \int_S \{|\delta\dot{y}_q|^2/v_q\}dx. \tag{10.4}$$

In view of Lemma 7.4 we have the further relations

$$\int_S |\delta y_q|\,|\delta\dot{y}_q|\,dx < ND(\delta y_q), \qquad \int_S |\delta y_q|^2 dx < ND(\delta y_q) \tag{10.5}$$

for a suitably chosen constant N.

We now define a sequence of variations $\{\eta_q\}$ by the formula

$$\eta_q = \delta y_q/k_q, \tag{10.6}$$

where k_q is the positive root of

$$k_q^2 = D(\delta y_q). \tag{10.7}$$

Using (10.4) and (10.5) it is seen that η_q satisfies the relations

$$\int_S \{|\dot{\eta}_q|^2/v_q\}dx = 1, \quad \int_S |\eta_q|\,|\dot{\eta}_q|\,dx < N, \quad \int_S |\eta_q|^2 dx < N. \tag{10.8}$$

By replacing our original subsequence $\{y_q\}$ by a suitable subsequence the variations η_q can be selected so that they have the additional properties described in the following

LEMMA 10.1. *The sequence of functions $\{y_q\}$ described above can be chosen so that the corresponding sequence of variations $\{\eta_q\}$ given by* (10.6) *converges almost everywhere to a variation η_0 in $\mathfrak{B}''$ and in fact is such that*

$$\lim_{q=\infty} \int_S |\eta_q - \eta_0|\,dx = 0, \qquad \lim_{q=\infty} \int_S |\eta_q^2 - \eta_0^2|\,dx = 0. \tag{10.9}$$

The functions η_q, $\dot{\eta}_{qk}$ $(k = 1, \cdots, m)$ converge weakly in $\mathfrak{L}$ to η_0, $\dot{\eta}_{0k}$

respectively. On every measurable subset T of S on which $\{\dot{y}_{qk}\}$ converges uniformly to $\dot{y}_{0k}$, the functions η_q, $\dot{\eta}_{qk}$ are in $\mathfrak{L}_2$ for large values of q and converge weakly in $\mathfrak{L}_2$ to η_0, $\dot{\eta}_{0k}$. Finally if $P(x)$, $Q_1(x), \cdots, Q_m(x)$ are continuous functions on the closure of S, we have

$$(10.10) \qquad \lim_{q=\infty} \int_S \{P\eta_q^2 + 2Q_i\eta_q\dot{\eta}_{qi}\}dx = \int_S \{P\eta_0^2 + 2Q_i\eta_0\dot{\eta}_{0i}\}dx.$$

In order to establish this result observe first that by virtue of the first relation in (10.8) and the inequality of Schwarz we have

$$\left| \int_M |\dot{\eta}_q|\, dx \right|^2 \leqq \int_M \{|\dot{\eta}_q|^2/v_q\}dx \int_M v_q dx \leqq \int_M v_q dx$$

for every measurable subset M of S. Moreover from the definition of v_q and $D(y)$ it follows that

$$\int_M v_q dx \leqq |M| + \int_M (v_q - 1)\,dx \leqq |M| + D(\delta y_q)$$

where $|M|$ is the measure of M. It follows from these relations together with (10.3) that the sequence of set functions

$$\int_M |\dot{\eta}_q|\, dx \qquad\qquad (q = 1, 2, 3, \cdots)$$

is uniformly bounded and uniformly absolutely continuous on S. By virtue of Theorem 8.3 our sequence can be replaced by a subsequence, again denoted $\{\eta_q\}$, that converges almost everywhere on S to a function η_0 in $\mathfrak{B}'$ and is such that the first equation in (10.9) holds. Moreover the functions η_q, $\dot{\eta}_{qk}$ converge weakly in $\mathfrak{L}$ to η_0, $\dot{\eta}_{0k}$ by the same theorem.

Consider now the function $z_{qk} = \dot{\eta}_{qk}/\sqrt{v_q}$. By virtue of the first relation in (10.8) it follows from Lemma 7.5 that by replacing our sequence by a suitable subsequence the corresponding functions z_{qk} will converge weakly in $\mathfrak{L}_2$ to a function z_{0k}. Since $\lim_{q=\infty} v_q = 1$ almost uniformly on S it follows that $z_{0k} = \dot{\eta}_{0k}$ almost everywhere on S and hence that η_0 is in the class $\mathfrak{B}''$. Moreover on each subset T of S on which $\lim_{q=\infty} v_q = 1$ uniformly the functions $\dot{\eta}_{qk}$ converge weakly to $\dot{\eta}_{0k}$ in $\mathfrak{L}_2$ since the functions z_{qk} have this property. From the last relation in (10.8) it follows that η_q converges weakly in $\mathfrak{L}_2$ to η_0 on S.

Consider next the function $\xi_q = \eta_q^2$ whose derivatives are given almost everywhere by the formula $\dot{\xi}_{qk} = 2\eta_q\dot{\eta}_{qk}$. By virtue of the last two equations in (10.8) we have

$$\int_S |\xi_q|\,dx < N, \qquad \int_S |\dot{\xi}_q|\,dx < 2N.$$

Consequently ξ_q is in $\mathfrak{B}'$. Using Theorem 8.2 it is seen that the second equation in (10.9) holds. This result together with Theorem 8.1 yields (10.10) with $\xi_q = \eta_q^2$, $\xi_0 = \eta_0^2$. This completes the proof of Lemma 10.1.

We proceed with the proof of Theorem 10.1 by establishing the following

LEMMA 10.2. *The Hilbert integral $J^*(y, c)$ of $J(y, c)$ given by formula (6.9) satisfies the relation*

$$(10.11) \qquad \lim_{q=\infty} \frac{J^*(y_q, c) - J^*(y_0, c)}{k_q^2} = \tfrac{1}{2}J^*_2(\eta_0, c)$$

for each element c in C_0. The variation η_0 is the one described in Lemma 10.1 and

$$(10.12) \qquad J^*_2(\eta, c) = \int_S \{g_{yy}\eta^2 + 2g_{yp_i}\eta\dot{\eta}_i\}\,dx,$$

the arguments in the derivatives of g being $[x, y_0(x), \dot{y}_0(x), c]$.

For by the use of Taylor's theorem it is seen by (6.9) that $J^*(y_q, c)$ can be written in the form

$$(10.13) \qquad J^*(y_q, c) = J^*(y_0, c) + J_1(\delta y_q, c) + \tfrac{1}{2}J^*_2(\delta y_q, c) + R_q(\delta y_q, c),$$

where $J_1(\eta, c)$ is given by (4.10) and $R_q(\eta, c)$ is of the form

$$R_q(\eta, c) = \int_S \{P_q\eta^2 + 2Q_{qi}\eta\dot{\eta}_i\}\,dx.$$

Since $\lim_{q=\infty} y_q = y_0$ uniformly on S an examination of the functions P_q, Q_{qi} will reveal that $\lim_{q=\infty} P_q = \lim_{q=\infty} Q_{qi} = 0$ uniformly for x on S and c in C_0. It follows from the last two relations in (10.8) that the first of the relations

$$\lim_{q=\infty} R_q(\eta_q, c) = 0 \qquad \lim_{q=\infty} J^*_2(\eta_q, c) = J^*_2(\eta_0, c) \qquad (c \text{ in } C_0)$$

holds. The second follows from equation (10.10). Using (4.10) we see

that $J_1(\delta y_q, c) = 0$. Using these relations together with (10.13) we obtain equation (10.11).

As was seen in **6** the function $J(y,c)$ can be written in the form

$$J(y,c) = J^*(y,c) + E^*(y,c) \tag{10.14}$$

where as in (5.3)

$$E^*(y,c) = \int_S E_g[x, y(x), \dot{y}_0(x), \dot{y}(x), c]\,dx.$$

We shall now prove the following

LEMMA 10.3. *The function $E^*(y,c)$ satisfies the relation*

$$\liminf_{q=\infty} [k_q^{-2} E^*(y_q, c)] \geqq \tfrac{1}{2} \int_S g_{p_i p_j} \dot{\eta}_{0i} \dot{\eta}_{0j}\,dx, \tag{10.15}$$

the argument in the derivatives of g being $[x, y_0(x), \dot{y}_0(x), c]$.

In order to establish this result let T be a subset of S on which $\dot{y}_{qk}$ converges uniformly to $\dot{y}_{0k}(x)$. Then $\lim_{q=\infty} v_q = 1$ uniformly on T and by (10.8) there exists an integer q_0 such that

$$\int_T |\dot{\eta}_q|^2 dx < 2 \qquad (q \geqq q_0). \tag{10.16}$$

Increase q_0 if necessary so that for $q \geqq q_0$ the functions

$$R_{qij}(x,c) = \int_0^1 g_{p_i p_j}[x, y_q(x), \dot{y}_0(x) + t(\dot{y}_q(x) - \dot{y}_0(x)), c]\,dt$$

are well defined on T. Then by Taylor's theorem

$$E_g(x, y_q(x), \dot{y}_0(x), \dot{y}_q(x), c) = \tfrac{1}{2} R_{qij} \delta\dot{y}_{qi} \delta\dot{y}_{qj}$$

whenever $q \geqq q_0$, x is in T and c is in C_0. Moreover, if we set

$$R_{ij}(x,c) = g_{p_i p_j}[x, y_0(x), \dot{y}_0(x), c]$$

we have $\lim_{q=\infty} R_{qij} = R_{ij}$ unformly for x on T. Consequently by (10.16)

$$\liminf_{q=\infty} \int_T R_{qij} \dot{\eta}_{qi} \dot{\eta}_{qj} = \liminf_{q=\infty} \int_T R_{ij} \dot{\eta}_{qi} \dot{\eta}_{qj}. \tag{10.17}$$

By an argument like that made in the proof of Lemma 9.1 we have

$$\liminf_{q=\infty} \int_T R_{ij} \dot{\eta}_{qi} \dot{\eta}_{qj}\,dx \geqq \int_T R_{ij} \dot{\eta}_{0i} \dot{\eta}_{0j}\,dx. \tag{10.18}$$

Observing that the integrand $E^*(y_q, c)$ is nonnegative for large values of q, by virtue of condition IIn on y_0, it follows that

$$\liminf_{q=\infty} k_q^{-2} E^*_{\,}(y_q, c) \geqq \liminf_{q=\infty} k_q^{-2} \int_T E_g(x, y_q, \dot{y}_0, \dot{y}_q, c)\, dx$$

$$\geqq \liminf_{q=\infty} \tfrac{1}{2} \int_T R_{qij} \dot{\eta}_{qi} \dot{\eta}_{qj}\, dx.$$

Combining this result with (10.17) and (10.18) we obtain the inequality

$$\liminf_{q=\infty} [k_q^{-2} E^*(y_q, c)] \geqq \tfrac{1}{2} \int_T R_{ij} \dot{\eta}_{0i} \dot{\eta}_{0j}\, dx.$$

Since T can be chosen so that the measure of $S - T$ is arbitrarily small this inequality also holds when $T = S$, as was to be proved.

LEMMA 10.4. *The variation η_0 is non null and satisfies the inequality $J_2(\eta_0, c) \leqq 0$ for every element c in C_0, where $J_2(\eta, c)$ is the second variation (4.14) of $J(y, c)$ on y_0.*

In order to establish this result observe that by (10.1)

$$(10.19) \qquad J(y_q, c) - J(y_0, c) < k_q^2/q.$$

Using the familiar relation

$$(10.20) \qquad J(y_q, c) - J(y_0, c) = J^*(y_q, c) - J^*(y_0, c) + E^*(y_q, c)$$

it follows from Lemma 10.2 that

$$(10.21) \qquad \liminf_{q=\infty} k_q^{-2}[J(y_q, c) - J(y_0, c)]$$
$$= \tfrac{1}{2} J^*{}_2(\eta_0, c) + \liminf_{q=\infty} (k_q^{-2} E^*(y_q, c)) \leqq 0.$$

From the definition of $J_2(\eta_0, c)$ and Lemma 10.3 we conclude that $J_2(\eta_0, c) \leqq 0$. If η_0 were null we would have $J^*{}_2(\eta_0, c) = 0$ and hence by (10.21)

$$\liminf_{q=\infty} k_q^{-2} E^*(y_0, c) \leqq 0.$$

On the other hand, the inequality (5.6) holds with $y = y_q$ if q is sufficiently large. It follows that there is a constant $h > 0$ such that

$$\liminf_{q=\infty} k_q^{-2} E^*(y_q, c) \geqq h > 0.$$

In view of this contradiction η_0 cannot be null. This proves Lemma 10.4.

The proof of Theorem 10.1 will be complete when we have established the following

LEMMA 10.5. *The variation η_0 is in the class $\mathfrak{B}_0$ described in* 2.

In order to prove this result it is sufficient to show that η_0 satisfies the equations (4.2). To this end observe that, by virtue of (10.19), (10.20), Lemma 10.2 and the nonnegativeness of $E^*(y_q, c)$, we have

$$(10.22) \qquad \lim_{q=\infty} k_q^{-1} E^*(y_q, c) = 0 \ .$$

Setting

$$J^*_\theta(y) = \int_C \{g_\theta(x, y(x), \dot{y}_0(x)) + (\dot{y}_i(x) - \dot{y}_{0i}(x)) g_{\theta p_i}(x, y(x), \dot{y}_0(x))\} dx$$

we have

$$J^*_\theta(y_0) = J_\theta(y_0)$$

and

$$J_\theta(y_q) - J_\theta(y_0) = J^*_\theta(y_q) - J^*(y_0) + E^*_\theta(y_q)$$

where $E^*_\theta(y, c)$ is given by (5.4). In view of (5.7) and (10.22) we have

$$\lim_{q=\infty} k_q^{-1} E^*_\theta(y_q) = 0 \ .$$

But by hypotheses $J_\theta(y_q) = J_\theta(y_0)$. Hence

$$(10.23) \qquad \lim_{q=\infty} k_q^{-1}[J^*_\theta(y_q) - J^*_\theta(y_0)] dx = 0.$$

Using Taylor's theorem again we see that

$$J^*_\theta(y) - J^*_\theta(y_0) = J_{\theta 1}(\delta y_q) + \int_S \{P_q \delta y_q + Q_{qi} \delta \dot{y}_{qi}\} dx$$

where $\lim_{q=\infty} P_q = \lim_{q=\infty} Q_{qi} = 0$ uniformly on S. Consequently

$$0 = \lim_{q=\infty} k_q^{-1}[J^*_\theta(y_q) - J^*_\theta(y_0)] = \lim J_{\theta 1}(\eta_q) = J_{\theta 1}(\eta_0)$$

as was to be proved.

11. Consequences of the sufficiency theorem. We begin with an extension of Theorem 6.2. To this end let $K(y, c)$ be a second integral having the properties described in the section preceding Theorem 6.2. Then we have the following

THEOREM 11.1. *Suppose that* y_0 *satisfies the conditions* I, IIn, III′, IV′ *relative to* $J(y,c)$. *Suppose further that* y_0 *satisfies condition* I *relative to* $K(y,c)$ *for every element* c *in* C *and that* (6.11) *holds as stated. There exists a compact subset* C_0 *of* C, *a constant* $\rho > 0$ *and a neighborhood* $\mathfrak{F}$ *of* y_0 *in xy-space such that given a function* y *in* $\mathfrak{A}_0$ *lying in* $\mathfrak{F}$ *the inequality*

$$J(y,c) - J(y_0,c) \geqq \rho \, | K(y,c) - K(y_0,c) | \tag{11.1}$$

holds for a suitably chosen element c *in* C_0.

In order to prove this result set

$$I(y,c,t) = J(y,c) - tK(y,c).$$

Since y_0 satisfies condition I relative to $J(y,c)$ and $K(y,c)$ it satisfies condition I relative to $I(y,c,t)$. Next choose C_0 as described in Theorem 9.1. From the proof of Theorem 6.2 it follows that there is a constant $\rho > 0$ such that y_0 satisfies condition IIn relative to $I(y,c,t)$ provided c is in C_0 and $|t| \leqq \rho$. By continuity considerations it follows that we can diminish ρ if necessary so that III′ also holds in this case. Finally diminish ρ so that it is effective as described in Theorem 9.2 when $K_2(\eta,c)$ is the second variation of $K(y,c)$ on y_0. The function $I(y,c,t)$ with $t = \pm\rho$ satisfies the hypothesis of Theorem 4.1. It follows that there is a neighborhood $\boldsymbol{F}$ of y_0 in xy-space such that the inequality

$$I(y,c,t) \geqq I(y_0,t,c) \qquad (t = \pm\rho)$$

holds for each function y in $\mathfrak{A}$ lying in $\boldsymbol{F}$ provided c is a suitably chosen element in C_0. The inequality (11.1) follows from this result.

There is a second consequence of Theorem 4.1 that we shall now describe. Let $[A,B]$ be an interval containing S in its interior. For a function η in $\mathfrak{B}'$ bounded by a constant δ we have, by Lemma 7.4,

$$\int_{A'_k}^{B'_k} |\eta(x_k,x'_k)|^2 dx'_k = \int_{A'_k}^{B'_k} \left| \int_{A_k}^{x_k} 2\eta(t_k,x'_k)\dot{\eta}_k(t_k,x'_k)\,dt \right| dx'_k$$
$$\leqq 2\int_A^B |\eta|\,|\dot{\eta}|\,dx \leqq ND(\eta),$$

where N is a suitably chosen constant. Let $D^*(\eta)$ denote the least number such that

$$\int_{A'_k}^{B'_k} |\eta(x_k,x'_k)|^2 dx'_k \leqq D^*(\eta) \qquad (k = 1,\cdots,m).$$

Then $D^*(\eta)$ is a second measure of the magnitude of η and

$$D^*(\eta) \leqq ND(\eta).$$

Consider now a function y in $\mathfrak{A}$ restricted to lie in the δ neighborhood of y_0. Then $D^*(y - y_0)$ is a second measure of the deviation of y from y_0. For the case $m = 1$ it takes the useful form

$$D^*(y - y_0) = \sup_{x \text{ on } S} |\, y(x) - y_0(x) \,|^2.$$

We have the following result

THEOREM 11.2. *The conclusion in Theorem* 4.1 *is valid when* $D(\eta)$ *is replaced by the function* $D^*(\eta)$ *described above.*

This theorem can be considered to be an extension of the Theorem of Osgood for simple integrals.

12. Sufficient conditions for a weak relative minimum. As is well known sufficient conditions for a weak relative minimum can be obtained from those for a strong relative minimum by a simple device that will be explained below. We shall describe this result for the problem formulated in **4**. It is a simple matter to interpret these results for the problem described in the introduction.

Let III$''$ and IV$''$ be the conditions obtained from the conditions III$'$ and IV$'$ described in **4** by omitting all references to the condition IIn. The weak sufficiency condition is then given by the following

THEOREM 12.1. *If the function* y_0 *satisfies the conditions* I, III$''$, IV$''$ *there exist a compact subset* C_0 *of* C, *a constant* $\rho > 0$ *and a neighborhood* R_0 *of* y_0 *in xyp-space such that given a function* y *in* $\mathfrak{A}_0$ *lying in* R_0 *the inequality*

$$J(y, c) - J(y_0, c) \geqq \rho D(y - y_0)$$

holds for a suitably chosen element c *in* C_0.

A function y in $\mathfrak{A}_0$ is said to lie in R_0 if its elements $(x, y(x), \dot{y}(x))$ are in R_0 for almost all values of x on S.

In order to prove this result choose C_0 as described in Theorem 9.1. By the use of a continuity argument and Taylor's theorem it is seen that if the region R described in **2** is diminished suitably the condition IIn will follow from condition III$''$ if we restrict the elements c to be in C_0. For this

new region R the conditions I, IIn, III′, IV′ will hold. It follows from Theorem 4.1 that ρ and R_0 can be chosen to have the properties described in Theorem 12.1, as was to be proved.

It is not difficult to see that for weak relative minima the function

$$D_1(\eta) = \int_S |\dot{\eta}|^2 dx$$

can be used in place of $D(\eta)$. This is not true in a strong relative minimum, except for special integrands $f(x, y, p)$.

Theorems analogous to Theorems 11.1 and 11.2 also hold for weak relative minima.

THE UNIVERSITY OF CHICAGO.

REFERENCES.

[1] S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932.

[2] G. A. Bliss, *The Calculus of Variations, Multiple Integrals.* Mimeographed notes of lectures delivered at the University of Chicago during the Spring quarter of 1939.

[3] G. A. Bliss, "The calculus of variations for multiple integrals," *American Mathematical Monthly*, vol. 49 (1942), pp. 77-89.

[4] J. W. Calkin, "Functions of several variables and absolute continuity, I," *Duke Mathematical Journal*, vol. 6 (1940), pp. 170-186.

[5] A. B. Carson, "An analogue of Green's theorem for multiple integral problems in the calculus of variations," *Dissertation*, University of Chicago, 1942.

[6] M. R. Hestenes, "The Weierstrass E-function in the calculus of variations," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 51-71.

[7] M. R. Hestenes, "The theorem of Lindeberg in the calculus of variations," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 72-92.

[8] M. R. Hestenes, "Sufficient conditions for the isoperimetric problem of Bolza in the calculus of variations," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 93-118.

[9] M. R. Hestenes, "An indirect sufficiency proof for the problem of Bolza in nonparametric form," *Transactions of the American Mathematical Society*, vol. 62 (1947), pp. 509-535.

[10] A. R. Jacoby, "Iterated integrals in the calculus of variations," *Dissertation*, University of Chicago, 1946.

[11] W. E. Karush, "A semi-strong minimum for a multiple integral problem in the calculus of variations," to appear soon in the *Transactions of the American Mathematical Society.*

[12] E. E. Levi, "Sui criterii sufficienti per il massimo e per il minimo nel Calcolo delle Variazioni," *Annali di Matematica Pura ed Applicata* (3), vol. 21 (1913), pp. 173-218.

[13] L. Lichtenstein, "Untersuchungen über zweidimensionale regulare Variationsprobleme," *Monatshefte für Mathematik und Physik*, vol. 28 (1917), pp. 3-51.

[14] E. J. McShane, "Sufficient conditions for a weak relative minimum in the calculus of variations," *Transactions of the American Mathematical Society*, vol. 52 (1942), pp. 344-379.

[15] C. B. Morrey, Jr., "Functions of several variables and absolute continuity, II," *Duke Mathematical Journal*, vol. 6 (1940), pp. 187-215.

[16] C. B. Morrey, Jr., "Multiple integral problems in the calculus of variations and related topics," *University of California Publications in Mathematics; New Series*: vol. 1 (1943), pp. 1-130.

[17] F. G. Myers, "Sufficient conditions for the problem of Lagrange," *Duke Mathematical Journal*, vol. 10 (1943), pp. 73-97.

[18] W. T. Reid, "Sufficient conditions by expansion methods for the problem of Bolza in the calculus of variations," *Annals of Mathematics*, vol. 38 (1937), pp. 662-678.

[19] W. T. Reid, "Sufficient conditions by expansion methods for multiple integral problems of the calculus of variations," *Bulletin of the American Mathematical Society*, vol. 45 (1939), p. 679.

ON THE REPRESENTATIONS OF THE SYMMETRIC GROUP.*

(Third Paper)

By G. DE B. ROBINSON.

Introduction. Perhaps the best known result of classical representation theory is that the number of conjugate sets is equal to the number of irreducible representations of a finite group G. In the case of the symmetric group S_n the association of the conjugate sets with the irreducible representations is made precise through the Young diagram $[\alpha]$ consisting of α_1 *nodes* in the first row, α_2 in the second, and so on, where

$$\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \cdots \geq \alpha_h,$$

and $\alpha_1 + \alpha_2 + \alpha_3 + \cdots + \alpha_h = n$. We shall refer to such a diagram as a *right* diagram, to distinguish it from a *skew* diagram $[\alpha] - [\beta]$, consisting of those nodes of $[\alpha]$ not belonging to $[\beta]$, with $\alpha_i \geq \beta_i$. These skew diagrams were studied in another paper [1] and it was shown that each such diagram which contains n nodes is associated with a reducible representation of S_n, whose irreducible components can be obtained according to a precise rule. The ideas had immediate application to yield a new proof of the Murnaghan-Nakayama recursion formula. We shall further investigate the properties of skew diagrams in Part I of the present paper, and apply the conclusions in Part II to obtain some new relations between the characters of the symmetric group.

Perhaps the most important results obtained in Part I have to do with skew diagrams whose disjoint constituents are right diagrams. Such diagrams arise when the q-hook structure [2] of a right diagram $[\alpha]$ is represented by means of a skew diagram $[\alpha]^{*}_{q}$ (see below). Also, these diagrams complete the story of the relationship between the symmetric group and the full linear group. For if $a_1 + a_2 + \cdots + a_r = n$, we may consider the irreducible Kronecker product representation

$$[a_1] \times [a_2] \times \cdots \times [a_r] \tag{1}$$

* Received April 18, 1947.

[1] [14]. The papers [13] and [14] will be referred to as SG_1 and SG_2.

[2] [11]. The papers [11] and [12] will be referred to as N_1 and N_2. Nakayama defines a p-hook to be any hook whose length is a power of p. Here we shall understand a q-hook to be a hook of length *exactly* q. The restriction that q be a prime will not apply here.

of the subgroup H, where H is the direct product

(2) $$S_{a_1} \times S_{a_2} \times \cdots \times S_{a_r}.$$

The representation (1) *induces* a representation of S_n which is characterized by a skew diagram whose disjoint constituents are the right diagrams $[a_i]$; such a representation may be denoted

(3) $$[a_1] \cdot [a_2] \cdot \cdots [a_r],$$

and *there is a one-to-one correspondence between the irreducible components of* (3) *and the irreducible components of the representation*

(4) $$\{a_1\} \times \{a_2\} \times \cdots \times \{a_r\}$$

of the full linear group.[3] The case where all the constituents $[a_i]$ in (3) are identical is important in the application of these ideas to invariant theory[4] which, it is hoped, will form the subject of a later paper.

Part II is largely devoted to a study of the Murnaghan-Nakayama recursion formula in the case where successive cycles of equal length q are removed from a given substitution. Let us assume that a maximum of b hooks of length q can be removed from a right diagram $[\alpha]$ containing n nodes, leaving a *q-core* $[\alpha_0]$ containing a nodes, so that

(5) $$n = a + bq;$$

here a may be zero. If fewer than b hooks are removed the residual diagram is not in general unique, but depends on the order of removing the hooks. On the other hand the q-core $[\alpha_0]$ is unique and is independent of the order in which the q-hooks are removed. In **6** it is proved that

(6) $$\chi_\alpha(V \cdot P_b) = \sigma \cdot x^*_q \cdot \chi_{\alpha_0}(V),$$

where χ_α is the character of $[\alpha]$, V is any substitution of the symmetric group S_a on the first a of the n symbols, P_b is the product of b cycles of length q on the remaining $n - a$ symbols, $\sigma = \pm 1$ and depends only on $[\alpha]$, and x^*_q is the number of distinct ways in which b hooks of length q can be removed from $[\alpha]$ to yield $[\alpha_0]$. If $V \cdot P_b$ contains *more* than b cycles of length q then its characteristic in $[\alpha]$ vanishes identically.

One can pursue the matter further by constructing a skew diagram $[\alpha]^*_q$ containing b nodes, each of which represents a particular q-hook of $[\alpha]$ and

---

[3] References to the literature will be found in SG_1; *cf.* in particular Murnaghan [10], p. 448 and [8]. The language here is different but the result is the same.

[4] Cf. [8] and [9].

such that every rq-hook of $[\alpha]$ is represented uniquely by an r-hook of $[\alpha]^*_q$. Clearly, the subgroup of S_n generated by those substitutions W of S_n on the last bq symbols which permute the b cycles of P_b is isomorphic to a symmetric group on b symbols which we may denote S^*_b. Actually, $[\alpha]^*_q$ yields a representation of S^*_b of degree x^*_q. The result (6) is generalized in 7 to read

$$(7) \qquad \chi_a(V \cdot P_b \cdot W) = \sigma \cdot \chi(W^*) \text{ in } [\alpha]^*_q \cdot \chi_{a_0}(V),$$

where W^* is the representative of W in S^*_b.

Apart from their intrinsic interest these results have an important application in the modular theory to the proof of Nakayama's conjecture [5] concerning the blocks of characters of S_n. The proof in question is based on a refinement of Nakayama's theorem concerning the prime factors of the degree x_a of $[\alpha]$. In its revised form the theorem reads

$$(8) \qquad e(x_a) = e(n!) - e((n-a)!) + e(x^*_p),$$

where $e(x_a)$ is the exponent of a given prime p dividing x_a, and here $n = a + bp$. The quantity $e(x^*_p)$ is the unknown ϵ, or *extra* power of p, appearing in Brauer's formula [6]

$$(9) \qquad e(x) = a - d + \epsilon,$$

where x is the degree of a given irreducible representation of an arbitrary group G, p^a is the exact power of p dividing the order g of G and p^d is the order of the appropriate *defect* group of G.

Since the writing of the present paper the author's attention has been drawn to a paper by Aitken [7] in which the skew diagram is introduced and, though the language used is that of Schur functions, the corresponding irreducible components are obtained. Aitken proves the result contained in Theorem 1. 3.

---

[5] N_2, p. 423. For the proof of the conjecture see [5] and [15]. For Nakayama's theorem concerning the prime factors of x_a see N_1, section 8.

[6] [4], p. 217.

[7] [1]. Recently, the papers left by Alfred Young have been made available to the author for study and it is of interest to note that Young had a very clear idea of a skew diagram as representing a *reducible representation* of S_n, though he says in his notes for lectures given in the tutumn of 1932 that he has not had an opportunity to develop the theory of such diagrams. (*Added in proof*, Feb. 12, 1948). In section III of Young's last paper (*Proceedings of the London Mathematical Society*, vol. 37 (1934), pp. 441-495) however, he does consider special skew diagrams obtained by shifting the upper rows of a right diagram to the right, and in section IV he obtains the irreducible components when there are just two rows. Further study of Young's unpublished MSS has brought to light a discussion of the irreducible components of a skew diagram consisting of two disjoint right diagrams. It is impossible to date these MSS accurately; Young died in December, 1940.

PART I.

Skew Representations.

1. In SG_2 we studied the reducible representation of the symmetric group S_n on n symbols which corresponds to the skew diagram $[\alpha(l)] - [\beta(m)]$, where $l = m + n$ and the notation $[\alpha(l)]$ serves to emphasize the fact that the right diagram $[\alpha]$ contains just l nodes. We shall change the notation slightly and denote the character of such a skew representation by $\chi_\alpha{}^\beta$ where χ_α is the character of the irreducible representation $[\alpha]$; we shall denote the degree of the reducible representation $[\alpha] - [\beta]$ by $x_\alpha{}^\beta$ and the degree of the irreducible representation $[\alpha]$ by x_α. It may happen that the skew diagram $[\alpha] - [\beta]$ breaks up into a number of *disjoint* constituents which have no row or column in common. If this is so, we may assume that the i-th constituent contains n_i nodes, where $n = n_1 + n_2 + \cdots + n_r$, and 4.11 of SG_2 becomes

1.1
$$x_\alpha{}^\beta = \frac{n!}{n_1!n_2!\cdots n_r!} x_{n_1}x_{n_2}\cdots x_{n_r},$$

where x_{n_i} denotes the degree of the corresponding representation of S_{n_i}, in general skew.

The symbol (n_i) will be used to denote a *skew* diagram containing n_i nodes, or the corresponding skew representation; the symbol $[n_i]$ will be used to denote a *right* diagram containing n_i nodes. When used as subscripts the brackets in these symbols will often be omitted if no ambiguity will result. Greek symbols will refer to particular right diagrams according to the usual custom.

If the rows and columns of a right diagram $[\alpha]$ are interchanged the resulting diagram $[\alpha]'$ is said to be *conjugate* to $[\alpha]$. If such a transformation is applied to the rows and columns of a skew diagram $[\alpha] - [\beta]$ then we obviously obtain $[\alpha]' - [\beta]'$, and the irreducible components of $[\alpha]' - [\beta]'$ are just the conjugates of those of $[\alpha] - [\beta]$. The following theorem is of greater significance:

1.2 *The irreducible representation* $[\alpha]$ *of* S_n *is similar to the skew representation* [8]

$$[\bar{\alpha}] - [\bar{\beta}] = [\alpha_1{}^h] - [\alpha_1 - \alpha_h, \cdots, \alpha_1 - \alpha_2, \alpha_1 - \alpha_1],$$

where $[\alpha_1{}^h]$ is used to designate the right diagram having h rows, each of

[8] The symbol $[\bar{\beta}]$ appears in D. E. Littlewood's work on invariants. Cf. [8], p. 316.

length α_1. Put in other words the theorem states that *the rotation of a right diagram through* 180° *is an invariant operation so far as the irreducible components of the representation of* S_n *are concerned.*

To prove 1.2 it is necessary to recall the method of obtaining the irreducible components of a skew representation as given in **3** of SG_2. Construct first the standard diagrams associated with the representation and the corresponding permutations; if any of these are lattice we have one of the irreducible components immediately. Transform the non-lattice permutations into lattice permutations, according to the *association* I of SG_1, and read off the irreducible component.

In the case under consideration it is evident that the process of rotating a standard right diagram D of $[\alpha]$ through 180° and reversing the assumed order of the symbols leads to a standard skew diagram $\bar{D}$ belonging to $[\bar{\alpha}] - [\bar{\beta}]$, and conversely; moreover, the permutation $p(D)$ associated with D is lattice while the permutation $p(\bar{D})$ is non-lattice. It will be sufficient to show that $p(\bar{D})$ is transformed into a lattice permutation $p(D_1)$ under the *association* I of SG_1, where D_1 ($\neq D$ in general) is a standard right diagram of $[\alpha]$. For this we remark that the *association* I can be reinterpreted in terms of the *association* II of SG_1, which leads to a definite operator *the same for every standard diagram* $[\alpha]$, which describes the raisings of symbols from the rows of $[\bar{\alpha}] - [\bar{\beta}]$ which are necessary to yield $[\alpha]$. Thus the non-lattice permutations obtained from the standard skew diagrams $[\bar{\alpha}] - [\bar{\beta}]$ are associated with the lattice permutations obtained from the standard right diagrams of $[\alpha]$, though not in the same order, which proves the theorem.

It is natural to generalize 1.2 to apply to the skew diagram $[\alpha] - [\beta]$. We shall suppose that the two right diagrams $[\alpha]$ and $[\beta]$ are taken as small as possible. This implies that nodes of $[\beta]$ appear in the first row of $[\alpha]$, and that the number of rows of $[\alpha]$ is greater than the number of rows of $[\beta]$; i. e.

$$[\alpha] = [\alpha_1, \alpha_2, \cdots \alpha_{h+k}], \qquad [\beta] = [\beta_1, \beta_2, \cdots \beta_h],$$

where $\alpha_i > \beta_i$ and $k > 0$. Then we have:

1.3 *The skew representation* $[\alpha] - [\beta]$ *of the symmetric group* S_n *is similar to the skew representation*

$$[\alpha_1, \cdots (k \text{ times}), \alpha_1 - \beta_h, \cdots, \alpha_1 - \beta_2, \alpha_1 - \beta_1] \\ - [\alpha_1 - \alpha_{h+k}, \cdots, \alpha_1 - \alpha_2, \alpha_1 - \alpha_1].$$

For brevity we may again denote the rotated skew diagram by $[\bar{\alpha}] - [\bar{\beta}]$.

The proof resembles that of 1. 2, with the added complication that one must take into consideration not only the operator required to change $[\bar{\alpha}]-[\bar{\beta}]$ into $[\alpha]-[\beta]$ but also that required to change the non-lattice permutations of $[\alpha]-[\beta]$ into lattice permutations. As before, these operators are the same for permutations associated with the same irreducible component; conversely, they are determined by the component.

To illustrate these ideas let $[\alpha]=[4,1]$, so that the corresponding right diagrams are as follows:

1. 4 $\quad$ 1234, 1235, 1245, 1345,
$\quad$ 5 $\quad$ 4 $\quad$ 3 $\quad$ 2

with the associated lattice permutations:

1. 5 $\quad c_1c_1c_1c_1c_2, \quad c_1c_1c_1c_2c_1 \quad c_1c_1c_2c_1c_1 \quad c_1c_2c_1c_1c_1.$

If the diagrams 1. 4 are rotated through 180° and the order of the symbols reversed, we obtain the four standard skew diagrams associated with the skew representations $[\bar{\alpha}]-[\bar{\beta}]=[4^2]-[3]$; these yield the non-lattice permutations:

1. 6 $\quad c_1c_2c_2c_2c_2, \quad c_2c_1c_2c_2c_2, \quad c_2c_2c_1c_2c_2, \quad c_2c_2c_2c_1c_2,$

which in turn lead to the lattice permutations 1. 5, but in reverse order.

2. Consider an arbitrary group G of finite order g and a subgroup H of order h. We can write

$$G=H+HS_2+HS_3+\cdots+HS_k,$$

and obtain a permutation representation of G on these k co-sets. The group matrix of the regular representation of G can be written:

2. 1
$$\begin{pmatrix} M & MS_2 & \cdots & MS_k \\ S_2^{-1}M & S_2^{-1}MS_2 & \cdots & S_2^{-1}MS_k \\ \cdot & \cdot & \cdots & \cdot \\ S_k^{-1}M & S_k^{-1}MS_2 & \cdots & S_k^{-1}MS_k \end{pmatrix},$$

where M is the group matrix $(x_{P^{-1}Q})$ of H and P, Q are any two elements of H. If M is replaced by any irreducible representation ω of H we obtain the representation of G *induced* [9] by ω.

We can now see the true significance of 4. 4 of SG_2, which, in the notation of this paper becomes

2. 2
$$(l!/m!)x_{(m)}=\Sigma x_a{}^{\beta}x_{a(l)}.$$

[9] [16], pp. 199-200.

If the left hand side of 2.2 be compared with 1.1, the suggestion is immediate that we have to do here with a skew diagram one of whose disjoint constituents is the right diagram $[\beta]$, the others being n isolated nodes. Such a skew diagram is associated with the representation of S_l induced by the irreducible Kronecker product representation

2.3 $$[\beta] \times [1]^n \sim [\beta]$$

of the sub-group S_m, where $[1]^n = [1] \times [1] \times \cdots$ (n factors). It is natural to denote the corresponding skew diagram by the symbol

2.4 $$[\beta] \cdot [1]^n.$$

A little consideration will show that every irreducible component of the skew representation 2.4 appears on the right hand side of 2.2 with the proper multiplicity, and only these. In fact, the process involved in deriving 2.2 is exactly the reverse of that required to find the irreducible components of 2.4. Thus we can write 2.2 in the form

2.5 $$[\beta] \cdot [1]^n = \Sigma x_\alpha{}^\beta [\alpha].$$

If this be compared with 4.3 of SG$_2$, i.e. with

2.6 $$[\alpha] = \Sigma x_\alpha{}^\beta [\beta],$$

attention being restricted to operations of S_m, we have an illustration of Frobenius' reciprocity theorem which states that the frequencies $x_\alpha{}^\beta$ must be the same in each case.

3. Thus we are led to consider those skew diagrams whose disjoint constituents are right diagrams. In the notation of the preceding paragraph, such a skew diagram would be denoted by

3.1 $$[n_1] \cdot [n_2] \cdot \cdots [n_r],$$

where the order of writing the constituents is unimportant. The sub-group H is here the direct product

3.2 $$S_{n_1} \times S_{n_2} \times \cdots \times S_{n_r},$$

and the representation of H under consideration is the Kronecker product

3.3 $$[n_1] \times [n_2] \times \cdots \times [n_r],$$

which is irreducible if and only if each of the representations $[n_i]$ is irreducible.[10] The degree of the induced representation 3.1 is given by 1.1.

[10] [6], p. 587.

Again we have a formula analogous to 2.5:

3.4 $$[n_1] \cdot [n_2] \cdot \cdots [n_r] \doteq \sum \lambda [n],$$

where λ is the number of ways in which the diagram $[n]$ can be built from the $[n_i]$ according to the Littlewood-Richardson rule. If there are just two constituents $[\beta]$ and $[\gamma]$, 3.4 becomes

3.5 $$[\beta] \cdot [\gamma] = \sum {}_\alpha\lambda_\beta{}^\gamma [\alpha],$$

which should be compared with 1.10 of SG_2, i. e. with

3.6 $$[\alpha] - [\beta] = \sum_\gamma {}_\alpha\lambda_\beta{}^\gamma [\gamma].$$

We state the conclusion in the form of a theorem:

3.7 *The frequency with which the representation* $[\gamma]$ *appears as an irreducible component of* $[\alpha] - [\beta]$ *is equal to the frequency with which* $[\alpha]$ *appears as an irreducible component of* $[\beta] \cdot [\gamma]$.

But one can go further and relate the general $(n_1) \cdot (n_2) \cdot \cdots (n_r)$ to $[n_1] \cdot [n_2] \cdot \cdots [n_r]$. This involves merely the repeated application of 3.6 to the constituents (n_i) to yield

3.8 $$(n_1) \cdot (n_2) \cdot \cdots (n_r) = \sum \mu [n_1] \cdot [n_2] \cdot \cdots [n_r].$$

We can combine 3.4 and 3.8 to yield an alternative and perhaps more natural proof of case (iii) of the Theorem proved in **6** of SG_2. Taking each of the (n_i) in 3.8 to be a skew hook we have what was there called an *incomplete* skew hook, and it follows from case (ii) of the Theorem that there is just one term on the right hand side of 3.8 in which each $[n_i]$ is a right hook; this is the only term which can contribute irreducible hook components $[n]$, according to 3.4. It is easily seen that the number of even hooks obtained in this way is equal to the number of odd ones. Clearly,[11] if one of the (n_i) is *not* a skew hook there can be no hook components $[n]$, which is case (i) of the Theorem.

One naturally asks the question, how are these skew representations related to the more familiar permutation representations? It is well known that the indentity representation must appear exactly once in any transitive permutation representation of a finite group.[12] This will happen in the case of a skew representation if and only if *each disjoint constituent consists of*

[11] Cf. [7], p. 107.
[12] [16], p. 181.

a single row. These permutation representations were denoted by the symbol $\Delta(\alpha)$ in SG_1.

The notation 3.1 suggests an important analogy with the Kronecker product[3]

3.10 $$\{n_1\} \times \{n_2\} \times \cdots \times \{n_r\}$$

of irreducible representations of the full linear group. We have, in fact, the theorem:

3.11 *There is a one-to-one correspondence between the irreducible components of the representation* 3.1 *of the symmetric group and the irreducible components of the representation* 3.10 *of the full linear group.*

4. We come now to the extension of the theory of hooks.[13] In N_1 Nakayama gives a careful analysis of the relation of one hook to another in a given right diagram $[\alpha]$, containing n nodes. Following Nakayama, we shall designate the right hook whose corner node is at the intersection of the i-th row and j-th column of $[\alpha]$ by the symbol (i, j), which symbol will also serve to define the equivalent skew hook consisting of those nodes in the corresponding part of the rim of $[\alpha]$. For two hooks (i, j), (s, t) of lengths g_1, g_2 respectively, there are three distinct cases:[14]

4.1
(i) $i \gtrless s,\ j \gtrless t$; one skew hook is completely contained in the other;
(ii) $\alpha_i \geqslant t$ or $\gamma_j \geqslant s$ with $i \gtrless s,\ j \lesseqgtr t$; the two skew hooks overlap;
(iii) $\alpha_i < t$ or $\gamma_j < s$ with $i \gtrless s,\ j \lesseqgtr t$; the two skew hooks have no common nodes.

Here the diagram $[\gamma] = [\alpha]'$, conjugate to $[\alpha]$. We shall have occasion to refer to this classification shortly. Nakayama's further conclusions which are significant here are summarized below.

4.2 *The removal of a hook of length $g = rq$ from a Young diagram $[\alpha]$ can always be accomplished*[15] *by the successive removal of r hooks of length q, and these q-hooks will not intersect in $[\alpha]$.*

If as many q-hooks as possible, say b, are removed from $[\alpha]$, then, either the

[13] It is important to notice here the shift in emphasis, as compared to N_1 and N_2, from the *right* hook to the *skew* hook.

[14] N_1, p. 169.

[15] N_1, section 5.

nodes of $[\alpha]$ are exhausted or a *q-core* $[\alpha_0]$ remains which contains a nodes and from which no further q-hooks can be removed.

4.3 *Let* $[\alpha_0]$ *be a q-core. For each* $r = 0, 1, 2, \cdots, q-1$, *there exists*[16] *one and only one diagram* $[\alpha]$ *such that the removal of a q-hook* $H_r = [q-r, 1^r]$ *from* $[\alpha]$ *reduces* $[\alpha]$ *to* $[\alpha_0]$.

Nakayama calls $[\alpha_0]$ the *kernel* of $[\alpha]$, but *q-core* is preferred here on account of its brevity, since the q must be made explicit, and because the word carries with it no other important mathematical connotation. It should be remarked that Nakayama insisted that q be a prime, but this is unnecessary at this stage and the assumption will not be made in the present paper.

The novelty in the following extension of the theory consists in determining *how* a succession of q-hooks can be removed from $[\alpha]$ to yield $[\alpha_0]$, and in particular, *in how many ways* these removals can be accomplished. We shall prove that *the q-hooks of* $[\alpha]$ *can be associated with the nodes of a new diagram* $[\alpha]^*_q$ *which completely represents the q-hook structure of* $[\alpha]$.

In the particular case where $[\alpha]$ is a right hook containing bq nodes our statement is obviously correct and $[\alpha]^*_q$ is a right hook containing b nodes. This implies that every cq-hook of $[\alpha]$ is represented by a c-hook of $[\alpha]^*_q$, and conversely. In the general case the situation is somewhat complicated and the reader may find it helpful in following the steps in the construction of $[\alpha]^*_q$ to refer to the example worked out at the end of the section.

We begin by locating the *longest* right hook in $[\alpha]$ of length c_1q. There may be several of the same length but fix attention upon one of them. This c_1q-hook H_1 of $[\alpha]$ gives rise to an equivalent skew hook $\bar{H}_1$ consisting of those nodes on the rim of $[\alpha]$ lying between the top right and bottom nodes of H_1, which nodes are of course the top and bottom nodes of $\bar{H}_1$. If the top node of $\bar{H}_1$ be taken to be (i, α_i), there will in general be ρ other skew hooks $\bar{H}_t$ of lengths c_tq, *all having the same top node* (i, α_i), and such that

4.4 $$\bar{H}_1 \supset \bar{H}_2 \supset \cdots \supset \bar{H}_\rho.$$

We may designate this sequence of skew q-hooks as the *q-chain* C in $[\alpha]$. If the corner node of H_1 is (i, j_1), and that of H_2 is (i, j_2) and so on, then it is evident that

4.5 $$j_1 < j_2 < \cdots < j_\rho.$$

But these conditions are precisely what are required to construct $[\alpha]^*_q$. We have in fact the equations:

[16] N_2, p. 414.

4.6
$$\begin{aligned}\gamma^*_1 &= c_1 - \rho + 1,\\ \gamma^*_2 &= c_2 - \rho + 2,\\ &\cdots\cdots\\ \gamma^*_\rho &= c_\rho - \rho + \rho,\end{aligned}$$

which completely determine the constituent $[\gamma^*]$ of $[\alpha]^*_q$, albeit in conjugate form.[17] As a matter of fact the coordinates of the c_i-hook H^*_i in $[\gamma^*]$ are just $(1, i)$.

Taking the longest hook in $[\alpha]$ whose length is a multiple of q and is not contained in C, proceed as before and construct a second q-chain C' and its corresponding constituent of $[\alpha]^*_q$, and so on until all the q-hooks of $[\alpha]$ have been included. In general, the $C^{(i)}$ will overlap as far as the nodes of the rim of $[\alpha]$ are concerned, but not so far as the q-hooks of $[\alpha]$ are concerned. In respect of these q-hooks the chains are separate entities. Thus we have the fundamental theorem:

4.7 *For any integer $q < n$ the q-hook structure of a right diagram can be completely represented by a skew diagram $[\alpha]^*_q$; each node of $[\alpha]^*_q$ represents a q-hook of $[\alpha]$ and each r-hook of $[\alpha]^*_q$ represents an rq-hook of $[\alpha]$.*

To each q-chain of $[\alpha]$ there corresponds a constituent of $[\alpha]^*_q$; these constituents are right diagrams in number at most q.

The final statement follows immediately from 4.3, since we may remove all but one q-hook from $[\alpha]$ and this hook can occupy at most q different positions which correspond respectively to top left hand corner nodes of the constituents of $[\alpha]^*_q$. Thus the number of these constituents is at most q.

To illustrate the above construction, consider the representation $[\alpha] = [8, 7, 4^2, 3, 2^2]$ of S_{30} and set $q = 2$. The longest hook $H_1 = (1, 1)$ is of length 7.2 and the top node of $\bar{H}_1$ is $(1, 8)$; starting from this node we also have $H_2 = (1, 3)$ of length 5.2, $H_3 = (1, 4)$ of length 4.2, and $H_4 = (1, 6)$ of length 2.2. Thus $\rho = 4$ for the chain C and $\gamma^*_1 = 4$, $\gamma^*_2 = 3 = \gamma^*_3$, and $\gamma^*_4 = 2$ as in the accompanying diagram 4.8. There remains a single hook $H' = (4, 2)$ of length 3.2 which has not been considered and this hook constitutes a second 2-chain C' of $[\alpha]$; there is no 2-core left. It follows that the second constituent of $[\alpha]^*_2$ must be the right diagram $[1^3]$.

---

[17] The equations 4.6 can also be written: $c_i = \gamma^*_i + \rho - i$, where $\rho = \alpha^*_1$, in which form they are already familiar as applied to $[\alpha]$ (cf. N_1, p. 167). It seems to the author that these equations take on new significance when thought of from this point of view.

$$4.8 \qquad [\alpha]:\ \begin{array}{cccccccc} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & \cdot & & & & \\ \cdot & \cdot & \cdot & & & & & \\ \cdot & \cdot & & & & & & \\ \cdot & \cdot & & & & & & \end{array} \qquad [\alpha]^*_2:\ \begin{array}{ccccc} & \cdot & \cdot & \cdot & \cdot \\ & \cdot & \cdot & \cdot & \cdot \\ & \cdot & \cdot & \cdot & \\ & \cdot & & & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \end{array}$$

There is one point in connection with the skew diagram $[\alpha]^*_q$ which requires further comment. As will appear from the preceding example, a given node of $[\alpha]$ may belong to several right 2-hooks depending on the order of removal of the hooks; this ambiguity arises from the existence of more than one constituent of $[\alpha]^*_2$. For if $[\alpha]^*_q$ had only one constituent, no two q-hooks of $[\alpha]$ corresponding to two nodes of $[\alpha]^*_q$ could intersect, by 4.2. On the other hand, if $[\alpha]^*_q$ has more than one constituent, then either (a) $[\alpha]-[\alpha_0]$ is connected and there exist skew hooks of $[\alpha]$ which overlap in the sense (ii) of 4.1 so that the corresponding right hooks will intersect, or (b) $[\alpha]-[\alpha_0]$ is not connected. In this latter case it may still be true that every node of $[\alpha]$ belongs to at most one right q-hook. We have, in fact, the following theorem:

4.9 *The necessary and sufficient condition that every node of $[\alpha]$ belong to at most one q-hook is that every disjoint constituent of $[\alpha]-[\alpha_0]$ be represented by a single constituent of $[\alpha]^*_q$.*

A little consideration will show that the operations ($'$), as in **1**, and (*) are permutable, i. e.

$$4.10 \qquad [\alpha]'^*_q = [\alpha]^{*\prime}_q.$$

PART II.

Characters.

5. In **6** of SG_2 we proved that[18]

$$5.1 \qquad \chi_\alpha^\beta(C) = (-1)^r \text{ or } 0,$$

according as $[\alpha]-[\beta]$ is a skew hook with n nodes equivalent to the right hook $H_r = [n-r, 1^r]$ or not, where C is a cycle on the n symbols. The equation 5.1 associates with the skew hook $\bar{H}_r$, equivalent to H_r, a *parity* $\sigma = (-1)^r = \pm 1$. We can state the following theorem:

[18] Cf. [10], p. 462.

5.2 *If at most b hooks of length q, of parities σ_i, can be removed from a right diagram $[\alpha]$ containing n nodes, where $n = a + bq$, then the quantity*

$$\sigma = \prod_{i=1}^{b} \sigma_i = (-1)^{\Sigma r_i} = \pm 1$$

is independent of the order of removal of the hooks.

To prove 5.2 we must utilize Nakayama's notion of the *image* of a hook.[19] Since all the hooks under consideration are of length q, case (i) of 4.1 does not concern us; in case (iii) there is no problem, so we can confine our attention to case (ii). Let us assume that

$$\gamma_j \geqslant s \text{ with } i < s,\ j > t.$$

If the q-hook (i, j) be removed from $[\alpha]$, we may denote the residual diagram by $[\alpha^{(1)}]$ and the image of (s, t) in $[\alpha^{(1)}]$ will be $(s-1, t)$. Conversely, if (s, t) be removed from $[\alpha]$ the image of (i, j) in the residual diagram $[\alpha^{(2)}]$ will be $(i, j-1)$. Now it is the sum of the lengths of the legs of these hooks which concerns us. Since the leg of the q-hook (u, v) is of length $\gamma_v - u$, we have

$$(\gamma_j - i) + (\gamma_t{}^{(1)} - (s-1)) = k_1, \qquad (\gamma_t - s) + (\gamma_{j-1}{}^{(2)} - i) = k_2,$$

where $k_1 = k_2 + 2$, since $\gamma_t{}^{(1)} = \gamma_t$ and $\gamma_{j-1}{}^{(2)} = \gamma_j - 1$. Thus k_1, k_2 are both even or both odd and the product of the corresponding σ_i's is constant. Successive applications of the argument prove the theorem, since all b hooks are removed in every case.

It will be important in the sequel to compare the parity of a cq-hook in $[\alpha]$ with that of its representative c-hook in $[\alpha]^*_q$, which necessarily belongs to a single constituent of $[\alpha]^*_q$. Let us take the cq-hook in question to be $H = [cq - r, 1^r]$ with parity $\sigma' = (-1)^r$ and suppose that its representative in $[\alpha]^*_q$ is $H^* = [c - s, 1^s]$. Then it is clear that

5.3 $$r = \sum_{i=1}^{c} r_i + s,$$

where $H_{r_i} = [q - r_i, 1^{r_i}]$ is the i-th of the c component q-hooks of H. This relation implies that

5.4 $$\sigma' = \sigma \cdot \sigma^*,$$

where $\sigma' = (-1)^r$, $\sigma = (-1)^{\Sigma r_i}$ and $\sigma^* = (-1)^s$.

[19] N_1, pp. 169-173.

6. Instead of considering the matrices of $[\alpha]$ associated with all the substitutions of S_n let us restrict attention to those of the sub-group

6.1 $$H = S_a \times S_q^{(1)} \times S_q^{(2)} \times \cdots \times S_q^{(b)},$$

where S_a is written on the symbols $1, 2, \cdots, a$, $S_q^{(1)}$ on the symbols $a+1, a+2, \cdots, a+q$, and so on. Such matrices yield a reducible representation of H, and if we denote any substitution of H by

6.2 $$U = V \times s^{(1)} \times s^{(2)} \times \cdots \times s^{(b)},$$

where V is any substitution of S_a and $s^{(i)}$ any substitution of $S_q^{(i)}$, we have the following relation between the characters:[20]

6.3 $$\chi_\alpha(U) = \Sigma \chi_{[a]}(V) \chi_{(q)}(s^{(1)}) \cdots \chi_{(q)}(s^{(b)}),$$

which corresponds to the removal of successive skew diagrams, each containing q nodes, from $[\alpha]$. Of course this process of removing skew diagrams could continue quite irrespective of the q-core $[\alpha_0]$ of $[\alpha]$, but we shall assume that it is repeated b times only. The significance of stopping at this stage is that some at least of those terms for which $[a] = [\alpha_0]$ involve q-hooks *only*. For $[a] \neq [\alpha_0]$ there must always be at least one $\chi_{(q)}$ which does not correspond to a skew q-hook. If the process were carried any further, *every* term would have at least one such factor.

We now simplify the situation drastically by assuming that each $s^{(i)}$ is a *cycle* on q symbols. It follows from 5.1 that *only those terms contribute to the right hand side of* 6.3 *in which every* (q) *is a skew hook, which implies that* $[a] = [\alpha_0]$. For any such term

6.4 $$\chi_{(q)}(s^{(1)}) \cdots \chi_{(q)}(s^{(b)}) = \sigma,$$

by 5.2. If we denote the number of such terms by x^*_q, then this is clearly also the number of standard skew diagrams $[\alpha]^*_q$. Thus we have proved the following theorem:

6.5 *If P_b represents the product of b cycles, each of length q, on the last bq of n symbols and V is any substitution of S_a on the first a symbols, then*

$$\chi_\alpha(V \cdot P_b) = \sigma \cdot x^*_q \cdot \chi_{\alpha_0}(V),$$

where $n = a + bq$.

If $P_{b'}$ contains b' cycles of length q, where $b' > b$ and $n = a' + b'q$, then *every* term on the right hand side of 6.3 will vanish in virtue of 5.1, since at most b hooks of length q can be removed from $[\alpha]$. Thus we have the following corollary to 6.5:

[20] SG_2, in particular Part II.

6.6 $$\chi_\alpha(V' \cdot P_{b'}) = 0,$$

for all V' of $S_{a'}$. It should be emphasized that *these results hold for all* $q < n$; the case $q = n$ is of little interest.

7. But we can go deeper still. Let us denote the normaliser of P_b in S_n by $\mathfrak{N}(P_b)$; then a little consideration will show that

$$\mathfrak{N}(P_b) \cong S_a \times P \times S^*_b,$$

where P is the sub-group generated by the b individual cycles of length q and S^*_b is isomorphic to the sub-group of substitutions W which permute the cycles of P_b amongst themselves. We prove the following generalization of Theorem 6.5:

7.1 *If W^* is the element of S^*_b which corresponds to W of S_n then*

$$\chi_\alpha(V \cdot P_b \cdot W) = \sigma \cdot \chi(W^*) \text{ in } [\alpha]^*_q \cdot \chi_{\alpha_0}(V).$$

Let us assume that the cycles in the product $P_b \cdot W = W \cdot P_b$ are of lengths $b_1q, b_2q, \cdots, b_kq$ where $b_1 + b_2 + \cdots + b_k = b$. Then we can write

7.2 $$\begin{aligned}\chi_\alpha(V \cdot P_b \cdot W) &= \Sigma\chi_{[a]}(V) \cdot \chi_{[a]}{}^{[a]}(P_b \cdot W) \\ &= \Sigma\chi_{[a]}(V) \cdot \chi_{(b_1q)}\chi_{(b_2q)} \cdots \chi_{(b_kq)}.\end{aligned}$$

As before, we are only interested in those terms which arise from a sequence of skew hook representations (b_iq), so that we can take $[a] = [\alpha_0]$. Consider one such term on the right hand side of 7.2 and apply 5.4 to its component hooks. Here σ is the product of the parities of the hooks of length q, σ' is the product of the parities of the b_iq-hooks, and σ^* is the product of the parities of the corresponding b_i-hooks in $[\alpha]^*_q$. I.e. multiplying the k equations 5.4 we obtain

7.3 $$\sigma' = \sigma \cdot \sigma^*.$$

It follows from 5.2 that σ is the same for all such terms of 7.2, so we can sum 7.3, applying the Murnaghan-Nakayama recursion formula iterated k times, to yield the equation

7.4 $$\chi_\alpha{}^{\alpha_0}(P_b \cdot W) = \sigma \cdot \chi(W^*) \text{ in } [\alpha]^*_q,$$

which proves 7.1.

As before, if $P_{b'}$ contains b' cycles of length q, where $b' > b$ and $n = a' + b'q$, we conclude that

7.5 $$\chi_\alpha(V' \cdot P_{b'} \cdot W') = 0,$$

for any V' of $S_{a'}$.

There is another approach to $[\alpha]^*_q$ which is of interest. We proved in the fundamental Theorem 4.7 that $[\alpha]^*$ is a skew diagram having at most q disjoint constituents, each of which is a right diagram. Hence we can write

7.6 $$[\alpha]^*_q = [b_1] \cdot [b_2] \cdot \cdots \cdot [b_q],$$

where $b_1 + b_2 + \cdots + b_q = b$, and we can think of the representation $[\alpha]^*_q$ as induced by an irreducible representation M of the sub-group $S_{b_1} \times S_{b_2} \times \cdots \times S_{b_q}$ as in **2** and **3**. Thus

7.7 $$\begin{aligned}\chi(W^*) \text{ in } [\alpha]^*_q &= x^*_q \text{ for } W^* = I,\\ &= \Sigma\Pi\chi_{b_i}(s_{b_i}) \text{ for } W^* = \Pi s_{b_i},\\ &= 0 \text{ otherwise.}\end{aligned}$$

These statements follow from the fact that the elements in the leading diagonal of 2.1 are M and its conjugates, which explains why a summation is required in the second line of 7.7.

As a simple illustration of these formulae consider the irreducible representation $[\alpha] = [4^2, 2^2]$ of S_{12} with $q = 2$. There is no 2-core, $[\alpha]^*_2 = [2,1] \cdot [2,1]$ and $\sigma = +1$. Take $P = (12)(34)(56)(78)(9\,10)(11\,12)$ and $W = (135)(246)$, $W^* = (abc)$, $P \cdot W = (145236)(78)(9\,10)(11\,12)$ and

7.8 $$\chi_\alpha(P) = + x^*_2 = 80, \qquad \chi_\alpha(P \cdot W) = + (-1)(2) \cdot 2 = -4,$$

since two terms in the diagonal of 2.1 must be taken into account.[21]

If we think of $[\alpha]^*_q$ as a sum of irreducible representations of S_b, then $\chi(W^*)$ becomes a sum of irreducible characters.

8. We shall conclude by deriving some interesting identities between the degrees of the irreducible representations of S_n. Consider the formula 2.5 with $n = bq$. If one were to suppose that no q-hooks could be removed from $[\beta]$ one might expect that, by introducing the product P_b of b cycles of length q and taking characters, all those $[\alpha]$ whose q-core is different from $[\beta]$ would be ruled out. Such a procedure would rule out those $[\alpha]$ whose q-core is larger than $[\beta]$, in view of 6.6, but it would not eliminate those $[\alpha]$ whose q-cores are on the same or a smaller number of nodes.

Consider the regular representation of S_n obtained by setting $[\beta] = 0$ in 2.5. In this case we have the familiar reduction

8.1 $$[1]^n = \sum_\alpha x_\alpha[\alpha],$$

where $n = bq$. Introducing P_b and taking characters we have

8.2 $$\chi(P_b) \text{ in } [1]^n = \sum_\alpha x_\alpha \sigma x^*_q,$$

[21] Cf. [17].

where the summation applies to all $[\alpha]$ which have no q-core; the other representations have been eliminated in virtue of 6.6. But the character on the left vanishes by 2.1, so that

8.3 $$\sum_a x_a \sigma x^*_q = 0.$$

On the other hand, one could introduce a substitution W which permutes the cycles of P_b and apply 7.1 to yield

8.4 $$\chi(P_b \cdot W) \text{ in } [1]^n = \sum_a x_a \sigma \chi(W^*) \text{ in } [\alpha]^*_q,$$

from which it follows as before that

8.5 $$\sum_a x_a \sigma \chi(W^*) \text{ in } [\alpha]^*_q = 0.$$

Once more the summation is over those $[\alpha]$ which have no q-core.

Choosing any irreducible representation $[b]$ of S_b we can multiply 8.5 by $\chi(W^*)$ in $[b]$, and sum over W^* to yield

8.6 $$\sum_a x_a \sigma \lambda = 0,$$

where λ is an integer $\geqslant 0$ which gives the multiplicity with which $[\alpha]^*_q$ contains the representation $[b]$ as an irreducible component.

We give as an illustration the case of $n = 6$ with $q = 2$. The following table, with the help of 4.10, gives the necessary information concerning σx^*_2 and the irreducible components of $[\alpha]^*_2$. Note that the representation

$[\alpha]$	$[6]$	$[5.1]$	$[4, 2]$	$[4, 1^2]$	$[3^2]$	$[3, 2, 1]$
x_a	1	5	9	10	5	16
$[\alpha]^*_2$	$[3]$	$[3]$	$[2] \cdot [1] =$ $[3] + [2, 1]$	$[2, 1]$	$[2] \cdot [1] =$ $[3] + [2, 1]$	—
σx^*_2	$+1$	-1	$+3$	-2	-3	—

$[3, 2, 1]$ is itself a 2-core and so will not appear. Corresponding to 8.3 we have the equation

8.7 $$x_6 - x_{5,1} + 3x_{4,2} - 2x_{4,1^2} - 3x_{3^2} + 3x_{2^3} + 2x_{3,1^3} - 3x_{2^2,1^2} + x_{2,1^4} - x_{1^6} = 0,$$

which is a consequence of the three equations:

$$
\begin{aligned}
&x_6 - x_{5,1} + x_{4,2} - x_{3^2} = 0,\\
8.8 \qquad &x_{4,2} - x_{4,1^2} - x_{3^2} + x_{2^3} + x_{3,1^3} - x_{2^2,1^2} = 0,\\
&x_{2^3} - x_{2^2,1^2} + x_{2,1^4} - x_{1^6} = 0,
\end{aligned}
$$

arising from 8.6 by setting $[b]$ equal to $[3]$, $[2,1]$, $[1^3]$ respectively. We have here a clear picture of the refinement resulting from the introduction of W or W^*.

UNIVERSITY OF TORONTO.

REFERENCES

[1] A. C. Aitken, "The monomial expansion of determinantal symmetric functions," *Proceedings of the Royal Society of Edinburgh* (A), vol. 61 (1943), pp. 300-310.

[2] R. Brauer, "On the arithmetic in a group ring," *Proceedings of the National Academy of Sciences*, vol. 30 (1944), pp. 109-114.

[3] ———, "On blocks of characters of groups of finite order" I, *ibid.*, vol. 32 (1946), pp. 182-186.

[4] ———, "On blocks of characters of groups of finite order" II, *ibid.*, vol. 32 (1946), pp. 215-219.

[5] ———, "On a conjecture by Nakayama," *Transactions of the Royal Society of Canada* (III), vol. 40 (1947), pp. 11-19.

[6] R. Brauer and C. Nesbitt, "On the modular characters of groups," *Annals of Mathematics*, vol. 42 (1941), pp. 556-590.

[7] D. E. Littlewood, *Group Characters*, Oxford, 1940.

[8] ———, "Invariant theory, tensors and group characters," *Philosophical Transactions of the Royal Society* (A), vol. 239 (1943), pp. 305-365.

[9] ———, "Invariants of systems of quadrics," *Proceedings of the London Mathematical Society*, vol. 49 (1947), pp. 282-306.

[10] F. D. Murnaghan, "On the representations of the symmetric group," *American Journal of Mathematics*, vol. 59 (1937), pp. 437-488.

[11] T. Nakayama, "On some modular properties of irreducible representations of a symmetric group," Part I, *Japanese Journal of Mathematics*, vol. 17 (1940), pp. 165-184.

[12] ———, Part II, *ibid.*, pp. 411-423.

[13] G. de B. Robinson, "On the representations of the symmetric group," *American Journal of Mathematics*, vol. 60 (1938), pp. 745-760.

[14] ———, Second Paper, *ibid.*, vol. 69 (1947), pp. 286-298.

[15] ———, "On a conjecture by Nakayama," *Transactions of the Royal Society of Canada* (III), vol. 40 (1947), pp. 20-25.

[16] A. Speiser, *Theorie der Gruppen von endlicher Ordnung*, 3rd ed., Berlin, 1937.

[17] M. Zia-ud-Din, "The characters of the symmetric group of degrees 12 and 13," *Proceedings of the London Mathematical Society*, vol. 42 (1937), pp. 340-355.

CRITERIA OF NON-DEGENERACY FOR THE WAVE EQUATION.*

By Philip Hartman and Aurel Wintner.

1. Let $f(s)$ be defined on an half-line $s >$ const., and let $x(s)$ denote an unspecified solution of the differential equation

$$x'' + f(s)x = 0. \tag{1}$$

It will always be understood that $f(s)$ is continuous, hence $x(s)$ has a continuous second derivative, and that $f(s)$ is required, hence $x(s)$ can be assumed, to be real-valued. Since $x(s)$ is continuous, such notions as its integrability will be meant to refer to its behavior as $s \to \infty$. In particular, $x(t)$ is of class (L^2) when

$$\int^{\infty} x^2(s)\,ds < \infty, \tag{2}$$

where the numerical value ($>$ const.) of the lower limit of integration is immaterial.

If α is a real constant and

$$f(s) = s^{\alpha}, \tag{3}$$

then (1) becomes a normal form of Bessel's equation. Correspondingly, it is easily verified from the asymptotic formula of the functions $J(s)$ that

(3*) according as the index α occuring in the case (3) of (1) does or does not exceed the value 2, every or no solution will satisfy (2),

provided that the trivial solution, $y(s) \equiv 0$, is not included in the second part of (3*). On the other hand, it is easily verified from the asymptotic formula of the functions $J(s)$, that the case (3) of (1) does or does not have a non-trivial solution $x(s)$ tending to 0, as $s \to \infty$, according as α does or does not exceed the value 0.

2. If two linearly independent, hence all, solutions of (1) are of class (L^2), then a (homogeneous) boundary condition assigned for the end of an s-half-line fails to define an eigenvalue problem for

$$x'' + (f(s) + \lambda)x = 0. \tag{4}$$

* Received October 27, 1947.

It is therefore important to delimit conditions which, *when imposed on* $f(s)$, preclude this degenerate case. The object of the first part of the present paper is to collect such *criteria of non-degeneracy.*

Corresponding criteria will then be obtained for the case in which (1) fails to possess any non-trivial solution of class (L^2). Since what is now required is that the number of linearly independent solutions of class (L^2), instead of being less than 2, be 0, the conditions to be imposed on $f(s)$ are now stricter than before; every criterion sufficient for an $f(s)$ of this type being automatically sufficient for an $f(s)$ of the former type. In other words, the eigenvalue problem is surely well-defined, and what is required of $f(s)$ is that $\lambda = 0$ be not in the point spectrum of (4), no matter how the boundary condition be chosen.

As proved in [2], this condition cannot be satisfied unless $\lambda = 0$ is either a point of the continuous spectrum or, with reference to every boundary condition, a cluster point of the point spectrum (possibly both). Since several of the sufficient criteria to be obtained happen to be such as to be satisfied by $f(s) + \text{Const.}$ whenever they are satisfied by $f(s)$, it follows that they contain sufficient criteria in order that *the entire line* $-\infty < \lambda < \infty$ *be in the continuous spectrum, while the point spectrum is vacuous.*

3. The simplest criterion of the *first* kind is this:

(i) *If* $f(s)$ *satisfies the unilateral restriction*

$$f(s) < \text{Const.}, \qquad (s \to \infty), \tag{5}$$

then (1) *cannot have two linearly independent solutions of class* (L^2).

Two proofs of (i) are known. The first is contained in the existence theories of A. Kneser ([5], § 2) and H. Weyl ([8], p. 238). The second (cf. [9], p. 8, end of § 3), being independent of existence proofs, reaches further; it concludes (i) as a corollary of the following elementary fact (which is known since a long time, but is not interesting, if (5) is replaced by $|f(s)| < \text{Const.}$; cf. [9], p. 6, end of § 1): If (5) is satisfied, then a solution of (1) cannot be of class (L^2) unless the derivative of the solution also is of class (L^2).

An adaptation of the second, rather than one of the first, proof of (i) will be needed in the proof of the following criterion:

(ii) *If* $f(s)$ *satisfies the unilateral restriction*

$$f(s_2) - f(s_1) < \text{Const.}\,(s_2 - s_1), \text{ where const.} < s_1 < s_2 < \infty, \tag{6}$$

(for instance, if $f(s)$ has a derivative satisfying

(6 bis) $$\limsup f'(s) < \infty, \qquad (s \to \infty),$$

while $\liminf f'(s) = -\infty$ *is allowed), then* (1) *cannot have two linearly independent solutions of class* (L^2).

As far as "orders of magnitude" are concerned, the following criterion is sharper than (ii);

(iii) *If $f(s)$ is monotone and satisfies the condition*

(7) $$\int^{\infty} |f(s)|^{-\frac{1}{2}}\, ds = \infty,$$

then (1) *cannot have two linearly independent solutions of class* (L^2).

In (iii), it can be asumed that

(8) $$df(s) \geqq 0$$

and

(9) $$f(s) > 0,$$

since otherwise the assertion of (iii) is contained in (i). By (8) is meant that the continuous function $f(s)$ is non-decreasing; its absolute continuity is not required. In other words, (8) should be interpreted as applying to estimations of Stieltjes integrals

$$\int_a^b p(s)\, df(s)$$

by the "first mean-value theorem."

Needless to say, (7), (8), (9) are neither necessary nor sufficient for (6); so that (iii) and (ii) are independent of each other. This implies that, even if only functions $f(s)$ satisfying (8) and (9) are admitted, (7) cannot be necessary for the truth of the assertion of (iii).

On the other hand, if only those functions $f(s)$ satisfying (8) and (9) are admitted which are of *regular growth*, then (7) is both necessary and sufficient for the truth of the assertion of (iii). This can be read off from a general asymptotic formula, deduced in [10] (cf. [3]) by a device of Liouville (or Riemann-Schwarz; cf. [7]). In this sense, the criterion supplied by (iii) is the best possible of its kind. The content of (3*) is the simplest manifestation of this situation.

4. *Proof of* (ii). Corresponding to the second proof of (i), which is a corollary of the fact that (1), (2) and (5) imply

$$\int^{\infty} x'^2(s)\,ds < \infty, \tag{11 bis}$$

(ii) will be proved as a corollary of the fact that (1), (2) and (6) imply

$$\limsup |x'(s)| < \infty \qquad (s \to \infty). \tag{11}$$

For, if the latter fact is granted, (iii) follows by considering the Wronskian of two linearly independent solutions of (1); cf. [9], p. 13.

On the other hand, (6) implies that the unilateral estimate

$$x'^2(s) + f(s)x^2(s) < \text{const.} \qquad (s \to \infty) \tag{12}$$

holds for every solution of (1) satisfying (2). In fact, if (6′) and (12′) denote the bilateral estimates which result if the functions occurring on the left of (6) and (12), respectively, are replaced by their absolute values, then (6′) implies (12′) if (2) is assumed. This was proved in [9], pp. 7-8. But a glance at the proof given *loc. cit.* makes it clear that (12′) can be replaced by (12) if (6′) is replaced by (6).

Accordingly, (ii) will be proved if it is shown that (12) implies (11).

If (1) is multiplied by x and the result is subtracted from (12), it follows that

$$x'^2(s) - x(s)x''(s) < \text{const.} \tag{13}$$

Two cases will be distinguished, according as $x''(s)$ is or is not distinct from 0 for every $s \geqq s^0$, if s^0 is sufficiently large.

In the first case, $x''(s)$ can be assumed to be positive on the half-line $s^0 \leqq s < \infty$ (in fact, $x(s)$ can be replaced by $-x(s)$ in (1), (2)). Then $x = x(s)$ is of class (L^2) on the half-line over which the (s, x)-graph is convex from above. This clearly implies that both $x(s)$ and $x'(s)$ tend to 0, as $s \to \infty$, which is more than what is claimed by (11).

In the second case, let μ denote the set of those points of the half-line $s^0 \leqq s < \infty$ at which $x''(s)$ vanishes. Since the half-line is closed and $x''(s)$ is a continuous function on it, μ is a closed set. By the assumption of the second case, μ contains points tending to ∞. On the other hand, (13) reduces to $x'^2(s) <$ const. on μ. Hence, (11) is true if s is restricted to μ. It remains to be shown that (11) holds without this restriction also.

Let $u_n < s < v_n$, where $n = 1, 2, \cdots$, be the sequence of open intervals which constitute the complement of the closed set μ with reference to the

closed half-line $s^0 \leqq s < \infty$. Since $x''(s)$ does not vanish on the complement of μ, the graph of $x = x(s)$ is either convex or concave on the closure of the n-th interval. Consequently, the maximum of $|x'(s)|$ for $u_n \leqq s \leqq v_n$ is attained either at $s = u_n$ or at $s = v_n$. Since u_n and v_n are in μ, and since (11) holds when s is restricted to μ, it follows that (11) remains true when s is not restricted to μ.

This completes the proof of (ii).

5. *Proof of* (iii). As mentioned after (iii), it can be assumed that, besides (7), the pair of supplementary conditions (8), (9) is satisfied. It will be shown that the function $x'/f^{\frac{1}{2}}$ of s must then be of class (L^2) whenever $x = x(s)$ is a solution of class (L^2). If this fact is granted, the assertion of (iii) results as a corollary, along the lines of the beginning of the proof of (ii), as follows:

Suppose that the assertion of (iii) is false, and denote by $x = x(s)$, $y = y(s)$ two linearly independent solutions of class (L^2). Since the Wronskian $xy' - yx'$ is a non-vanishing constant, it follows from (7) and (9) that $f^{-\frac{1}{2}}$ times this Wronskian cannot be of class $(L) = (L^1)$. But this is impossible if $x'/f^{\frac{1}{2}}$, $y'/f^{\frac{1}{2}}$ must be of class (L^2) whenever x, y are of class (L^2). In fact, since the product of two functions of class (L^2) is of class (L), it is seen that $f^{-\frac{1}{2}}$ times the Wronskian becomes the difference of two functions of class (L) and is therefore of class (L).

Accordingly, it is sufficient to prove that $x'/f^{\frac{1}{2}}$ must be of class (L^2) if x is of class (L^2); in other words, that (2) implies

$$\int^{\infty} \{x^2(s) + x'^2(s)/f(s)\} ds < \infty. \tag{14}$$

6. It turns out that the truth of this implication is independent of the assumption (7) of (iii). In fact, the following Lemma involves only (8) and (9).

LEMMA. *If* $f(s)$ *is positive and non-decreasing, then every solution of* (1) *satisfying* (2) *satisfies* (14) *also.*

In order to prove this, a few standard facts will have to be collected.

First, if $f(s)$ is any continuous function, then

$$\int_a^b \{x'^2(s) + f(s)x^2(s)\} ds = \int_a^b x^2(s) df(s) \tag{15*}$$

is an identity in a, b along every solution $x = x(s)$ of (1); cf. [9], pp. 7-8. If, in addition, $f(s) \neq 0$ for $a \leqq s \leqq b$, then

$$(16^*) \qquad \int_a^b \{x^2(s) + x'^2(s)/f(s)\} ds = \int_a^b x'^2(s) d\{1/f(s)\}$$

can be verified in exactly the same way as (15*). If (8) is assumed, it follows from (15*) that

$$(15) \qquad d\{x'^2(s) + f(s)x^2(s)\} \geqq 0,$$

and from (16*) that

$$(16) \qquad d\{x^2(s) + x'^2(s)/f(s)\} \leqq 0,$$

provided that $f(s) \neq 0$, *e. g.*, that (9) is assumed (the inequalities in (15), (16) are meant in the same sense as in (9); so that $f(s)$ need not have absolute continuity).

Next, (8) and (9) imply that, if $s^0 \leqq s < \infty$ and $c^2 = f(s^0)$, then (1) is minorized, in Sturm's sense, by the linear oscillator $y'' + c^2 y = 0$, where $c \neq 0$. Hence, if $x(s)$ is any non-trivial solution of (1), then $x(s)$ has an infinite sequence of isolated zeros (which tend to ∞). Let $s_0 < s_1 < s_2 < \cdots$ be the sequence of all these zeros. Then, since (9) implies that $u'' + f(s + s_n)u = 0$ is minorized, in Sturm's sense, by $v'' + f(s + s_{n+1})v = 0$,

$$(17) \qquad s_n - s_{n-1} \geqq s_{n+1} - s_n.$$

Since $x(s)$ can be replaced by the solution $-x(s)$, there is no loss of generality in assuming that $x(s)$ is positive between s_0 and s_1. Then, since a zero of a non-trivial solution of (1) cannot be a multiple zero,

$$(18) \qquad (-1)^n x(s) > 0 \text{ when } s_n < s < s_{n+1}, \qquad (x(s_n) = 0).$$

Since (1) and (9) imply that $x(s)$ and $x''(s)$ cannot have opposite signs, i. e., that the curve $x = x(s)$ must turn its concavities toward the s-axis, it is seen from (18) that the graph of $y = |x(s)|$ is a convex arch over each of the intervals $s_{n-1} \leqq s \leqq s_n$.

In particular, there exists on the interval $s_{n-1} < s < s_n$ a unique point s, say $s = s_n^*$, at which $x'(s)$ vanishes. It is also clear from the convexity of the arch that

$$(19) \qquad |x(s)| \geqq \tfrac{1}{2} x(s_n^*) \text{ when } p_n \leqq s \leqq q_n,$$

where $p_n \leqq s \leqq q_n$ is a certain interval contained in the interval $s_{n-1} \leqq s \leqq s_n$ and having a length subject to the inequality

(19′) $$q_n - p_n \geqq \tfrac{1}{2}(s_n - s_{n-1}).$$

On the other hand, since s_n^*, s_{n+1}^* are within the respective s-intervals (s_{n-1}, s_n), (s_n, s_{n+1}),

$$s_{n+1} - s_{n-1} > s_{n+1}^* - s_n^*.$$

Finally, (17) can be written in the form

$$s_n - s_{n-1} \geqq \tfrac{1}{2}(s_{n+1} - s_{n-1}).$$

Since the last three formula lines imply that

$$q_n - p_n > (s_{n+1}^* - s_n^*)/4,$$

it follows from (19) that

$$\int_{p_n}^{q_n} x^2(s)\,ds > x^2(s_n^*)\,(s_{n+1}^* - s_n^*)/16.$$

But the intervals (p_1, q_1), (p_2, q_2), $\cdots$ are disjoint. Hence, the last inequality implies that

$$\sum^{\infty} x^2(s_n^*)\,(s_{n+1}^* - s_n^*) < \infty \text{ if } \int^{\infty} x^2(s)\,ds < \infty.$$

On the other hand, since s_n^* has been defined by $x'(s_n^*) = 0$, it is clear from (16) alone that

$$x^2(s_n^*) \geqq g(s) \text{ for } s_n^* \leqq s \leqq s_{n+1}^*,$$

if $g = g(s)$ is an abbreviation for the (positive) function $x^2 + (x'^2/f)$ of s.

The last two formula lines imply that

$$\sum^{\infty} G(s_n^*, s_{n+1}^*) < \infty \text{ if } \int^{\infty} x^2(s)\,ds < \infty,$$

where

$$G(a, b) = \int_a^b g(s)\,ds; \qquad g = x'^2 + (x^2/f).$$

Since $s_n^* \to \infty$ as $n \to \infty$, it follows that (14) must be true if (2) is assumed. This proves the Lemma. Hence, the proof of (iii) is now complete.

7. The following variant of (iii) will now be proved:

(iv) *If $f(s)$ has a derivative satisfying*

(20) $$\limsup |f'|/|f|^{\frac{3}{2}} < \infty, \qquad (s \to \infty),$$

6

(so that, in particular, $f(s) \neq 0$ when s is large), and if

(7) $$\int^{\infty} |f(s)|^{-\frac{1}{2}} ds = \infty,$$

then (1) *cannot have two linearly independent solutions of class* (L^2).

For another application of (20), cf. [1].

It is clear from the proof of (iii) that (iv) is a corollary of the following variant of the Lemma:

If $f(s)$ is positive and has a derivative satisfying (20), *then every solution of* (1) *satisfying* (2) *satisfies* (14) *also.*

In order to prove this variant of the Lemma, multiply (1) by x and use the identity $xx'' = (xx')' - x'^2$. This gives

$$(xx')' - x'^2 + fx^2 = 0$$

or, since $f > 0$,

$$(xx')'/f - x'^2/f + x^2 = 0.$$

It follows therefore from (2) that the integral

$$\int^{r} \{(xx')'/f - x'^2/f\} ds$$

tends to a finite limit, as $r \rightarrow \infty$. Hence, a partial integration of the first term of this integral shows that the expression

$$x(r)x'(r)/f(r) + \int^{r} (xx'f'/f^2 - x'^2/f) ds$$

tends to a finite limit, and remains therefore bounded, as $r \rightarrow \infty$.

Next, every solution $x(s)$ of (1) must have a sequence of zeros s tending to ∞. For otherwise it can be assumed that, as $s \rightarrow \infty$, the value of $x(s)$ is ultimately positive. Then, since $f(s) > 0$ by assumption, it follows from (1) that $x''(s)$ is ultimately negative. Consequently, $x'(s)$ is ultimately positive. Accordingly, $x(s)$ is positive and increasing from a certain s onward. But this contradicts (2).

Hence, r can tend to ∞ through a sequence of values $s = r$ at which $x(s) = 0$. Since the expression in the last formula line was seen to be bounded, it follows that

(21) $$\int^{r} (xx'f'/f^2 - x'^2/f) ds$$

is bounded on a sequence of r-values tending to ∞.

On the other hand, if the inequality $|ab| \leqq \frac{1}{2}(a^2 + b^2)$ is applied to $a^2 = x'^2/f$ and $b^2 = x^2f'^2/f^3$, it is seen that the absolute value of (21) is minorized by

$$\tfrac{1}{2}\int^r (x'^2/f - x^2f'^2/f^3)\,ds$$

for every r. This means that

$$\int^r (x'^2/f)\,ds \tag{22}$$

is majorized by 2 times the absolute value of (21) plus

$$\int^r (x^2f'^2/f^3)\,ds.$$

But (2) and (20) imply that the last integral remains bounded as $r \to \infty$. Since (21) remains bounded when r tends to ∞ on a certain sequence, it now follows that the same is true of (22). But (22) is a non-decreasing function of r, since $f > 0$. Hence, (22) remains bounded when r tends to ∞ continuously. In view of (2), this proves (14).

The proof of the last italicized statement, and therefore that of (iv), is now complete.

8. Those criteria will now be considered which are of the *second* type (in the sense of the description given between (4) and (5) above). A criterion of this type can be formulated as follows:

(I) *If $f(s)$ is positive, monotone and such as to satisfy*

$$\int^\infty \{f(s)\}^{-1}ds = \infty, \tag{23}$$

then (1) *has no non-trivial solution* ($\not\equiv 0$) *of class* (L^2).

Needless to say, (iii) assumes less, but claims less, than (I). On the other hand, the above criteria of the first kind, *viz.*, (i), (ii), (iii) and (iv), fail to contain even that criterion of the first kind which is a corollary of the following criterion of the second kind:

(II) *If $f(s)$ is positive, monotone and such as to satisfy*

$$\limsup f(s)/s < \infty, \qquad (s \to \infty), \tag{24}$$

then (1) *has no non-trivial solution* ($\not\equiv 0$) *of class* (L^2).

Whereas (24) is necessary, though not sufficient, for the assumption (6) of (ii), the assumption of (II) requiring the monotony of $f(s)$ is not stipulated in (ii).

Actually, (24) implies (23); so that (II) is contained in (I). On the other hand, the following variant of (II) is not implied by (I):

(III) *If $f(s)$ is positive, monotone and such that there exists a sequence $r_1, r_2, \cdots$ satisfying*

$$\limsup f(r_n)/r_n < \infty \text{ and } r_n \to \infty, \qquad (n \to \infty), \tag{25}$$

then (1) *has no non-trivial solution* ($\not\equiv 0$) *of class* (L^2).

9. *Proof of* (I). Let $x = x(s)$ be an arbitrary non-trivial solution of (1). It must be shown that x does not satisfy (2).

First, the argument employed before (21) shows that, if the zeros of $x(s)$ do not cluster at ∞, then $x(s)$ is not of class (L^2). Hence, it can be assumed that $x(s)$ has an infinite sequence of isolated zeros (which tend to ∞). Let $s_0 < s_1 < s_2 < \cdots$ be the sequence of all these zeros. The description of the graph of $y = |x(s)|$, given after (17), is applicable in the present case, as are (19) and (19′).

Since f is monotone, it is either non-increasing or non-decreasing. Suppose first that f is non-increasing. It then follows from (16*) that

$$g(s) = x^2(s) + \{x'^2(s)/f(s)\} \tag{26}$$

is non-decreasing. Hence, since $s_1^* < s_2^* < \cdots$ and $x'(s_n^*) = 0$,

$$0 < x^2(s_1^*) \leqq x^2(s_2^*) \leqq \cdots.$$

Consequently, from (19) and (19′),

$$\sum_{n=1}^{\infty} I(p_n, q_n) \geqq (\tfrac{1}{2})^2 x^2(s_1^*) \sum_{n=1}^{\infty} (s_n - s_{n-1}),$$

where

$$I(a, b) = \int_a^b x^2(s)\,ds.$$

Since the intervals $(p_1, q_1), (p_2, q_2), \cdots$ are disjoint, $x^2(s_1^*)$ is positive, and $s_n \to \infty$ as $n \to \infty$, it follows that (2) cannot hold.

Next, suppose that $f(s)$ is non-decreasing. Then (15) holds. But (15) and (9) imply that, since the trivial solution $x(s) \equiv 0$ has been excluded, there exists a positive constant c satisfying

$$x'^2 + fx^2 > c > 0$$

as $s \to \infty$. Accordingly, if s is large enough,

$$x^2 + (x'^2/f) > c/f > 0. \tag{27}$$

It follows therefore from (23) that (14) cannot hold, and so, since the conditions of the Lemma are fulfilled, (2) cannot be satisfied either.

This completes the proof of (I).

10. *Proof of* (III). It can be assumed that $f(s)$ is non-decreasing (otherwise (III) is contained in (I)). Then, if $x(s)$ is any non-trivial solution, (27) is applicable. In addition, (8) and (9) make applicable all notations and formulae occurring in the proof of the Lemma. Furthermore, (8) implies (15), and so the function (26) is non-increasing.

Since $s_1^* < s_2^* < \cdots$ and $x'(s_n^*) = 0$,

$$x^2(s_1^*) \geqq x^2(s_2^*) \geqq \cdots.$$

It follows therefore from (19) that the inequality

$$x^2(s) \geqq (\tfrac{1}{2})^2 x^2(s_n^*)$$

holds at every s contained in any of the n intervals $(p_1, q_1), (p_2, q_2), \cdots (p_n, q_n)$. But the last inequality implies, by (27), that

$$x^2(s) > (\tfrac{1}{2})^2 c/f(s_n^*),$$

since $x'(s_n^*) = 0$. Hence, by the monotony of f,

$$x^2(s) > (\tfrac{1}{2})^2 c/f(r_k)$$

holds if $s_n^* \leqq r_k$. Consequently, by (19′),

$$\int_{s_j}^{s_n} x^2(s)\,ds > \tfrac{1}{2}(s_n - s_j)(\tfrac{1}{2})^2 c/f(r_k) \text{ if } s_j < s_n \leqq r_k.$$

On the other hand, the assumptions (8) and (9) imply (17); in particular, $s_1 - s_0, s_2 - s_1, \cdots$ is a bounded sequence. Hence, if s_n is the largest zero of $x(s)$ not exceeding r_k, it follows from the second of the assumptions in (25) that

$$s_n/r_k \to 1 \text{ as } k \to \infty.$$

If j is fixed but k (and so n) tends to ∞, the last two formula lines supply the estimate

$$\int_{s_j}^{\infty} x^2(s)\,ds \geqq (\tfrac{1}{2})^3 c \liminf r_k/f(r_k).$$

The first assumption in (25) then shows that there exists a positive constant C satisfying

$$\int_{s_j}^{\infty} x^2(s)\,ds > C > 0.$$

But the existence of such a $C = \text{const.}$ contradicts (2), since $s_j \to \infty$ as $j \to \infty$. This contradiction proves (III).

11. The remarks, preceding the beginning of the proof of (ii), concerning the necessity (rather than the sufficiency) of (7) in (iii) when f is of a regular growth, can be paralleled (and made more precise) when (iii), a criterion of the first kind, is replaced by (I), a criterion of the second kind. In fact, the situation is as follows:

(IV) *If $f(s)$ is of "regular growth" and tends to ∞ as $s \to \infty$, then* (7) *is necessary and sufficient in order that* (1) *should not possess any non-trivial solution of class* (L^2).

On the other hand, if $f(s)$ tends increasingly to ∞ without being of "regular growth," then, while (23) *is still sufficient,* (7), *and even*

$$\int^{\infty} \{f(s)\}^{-1+\epsilon}\,ds = \infty \text{ for every } \epsilon > 0, \tag{28}$$

fails to be necessary.

There is no point in making here precise that meaning of the "regularity" of growth under which the first assertion of (IV) is true. Suffice it to mention that if $f(s)$ is an "L-function" (*i. e.*, a logarithmic-exponential function), then, in order that (1) should have no non-trivial solution of class (L^2), condition (7) is both necessary and sufficient. This can be read off from the explicit asymptotic results referred to before the beginning of the proof of (ii).

Since (I) asserts the sufficiency of (23) without a condition of regularity, it follows that the proof of (IV) will be complete if it is ascertained that there exists an increasing function $f(s)$ which tends to ∞ with s and satisfies (28), although (1) has a non-trivial solution satisfying (2).

If $f(s)$ is allowed to be a step-function (hence discontinuous), an hypothesis which simplifies the construction but is hardly material, then an adaptation of a construction of Milloux [6], p. 50, leads to examples of the desired type, as follows:

In terms of a sequence of numbers $A_1, A_2, \cdots$ satisfying

$$0 < A_1 < A_2 < \cdots < A_n \to \infty$$

and

$$a_n \to \infty, \text{ where } a_n = 2\pi \sum_{k=1}^{n} A_k^{-1},$$

define $f(s)$ by placing

$$f(s) = A_n^2 \text{ for } a_{n-1} \leqq s < a_n.$$

Then $f(s)$ satisfies (8), (9) and, since

$$\int^{\infty} \{f(s)\}^{-1+\epsilon}\, ds = 2\pi \sum^{\infty} A_n^{-3+2\epsilon},$$

(28) becomes satisfied by choosing, for instance,

$$A_n^3 = n \log^2 (n+1).$$

Clearly, this choice of A_n satisfies all of the above requirements and is such as to make the series

$$\sum_{n=1}^{\infty} A_n^{-3}$$

convergent. Hence, the proof will be complete if it is verified that a constant multiple of the latter series is a majorant of the integral (2) belonging to a non-trivial solution of (1).

The definition shows that (1) reduces to

$$x'' + A_n^2 x = 0 \text{ for } a_{n-1} \leqq s < a_n$$

and admits, therefore, the solution

$$x(s) = A_n^{-1} \sin \{A_n(s - \alpha_n)\} \text{ for } a_{n-1} \leqq s < a_n.$$

Since $a_n - a_{n-1} = 2\pi A_n^{-1}$, the choice $\alpha_n = a_{n-1}$ of the arbitrary phases α_n lets $x(s)$ acquire a continuous second derivative for every s. On the other hand, the last formula line implies that

$$\int_{a_{n-1}}^{a_n} x^2(s)\,ds = A_n^{-3} \int_0^{2\pi} \sin^2 s\,ds.$$

This proves that the integral (2) is majorized by a constant multiple of ΣA_n^{-3}.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

[1] P. Hartman and A. Wintner, "The asymptotic arcus variation of solutions of linear differential equations of second order," *American Journal of Mathematics*, vol. 70 (1948), pp. 1-10.

[2] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[3] G. H. Hardy, "Orders of infinity" (Cambridge Tracts, no. 12), Cambridge, 1924 (2nd ed.), pp. 33-37.

[4] F. Klein, "*Vorlesungen über die hypergeometrische Funktion*" (ed. O. Haupt), Berlin, 1933.

[5] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser linearer Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

[6] H. Milloux, "Sur l'équation différentielle $x'' + xA(t) = 0$", *Prace Matematyczno-Fizyczne*, vol. 41 (1934), pp. 39-54.

[7] B. Riemann, "*Gesammelte mathematische Werke*," Leipzig, 1892 (2nd ed.), pp. 335-336.

[8] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[9] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

[10] ———, "On the normalization of the characteristic differentials in continuous spectra," *Physical Review*, vol. 72 (1947), pp. 516-517.

ON THE ORIENTATION OF UNILATERAL SPECTRA.*

By PHILIP HARTMAN and AUREL WINTNER.

Consider the differential equation

(1) $$(px')' + qx = 0$$

on the half-line $0 \leqq s < \infty$, on which the coefficient functions

(2) $$p = p(s) > 0 \text{ and } q = q(s) \gtreqless 0, \text{ where } 0 \leqq s < \infty,$$

are continuous (but some or all of the four signs of equality occurring in

$$0 \leqq \liminf_{s\to\infty} p, \quad \limsup_{s\to\infty} p \leqq \infty, \quad -\infty \leqq \liminf_{s\to\infty} q, \quad \limsup_{s\to\infty} q \leqq \infty$$

are allowed), and suppose that (1) does not possess two linearly independent solutions of class (L^2). By this is meant that at least one solution $x = x(s)$ of (1) is not in Hilbert's space

(3) $$\int^{\infty} x^2(s)\,ds < \infty.$$

The formulation (3) of the (L^2)-condition is justified because, on the one hand, only real-valued solutions of (1) need to (and will) be considered and, on the other hand, the lower limit of integration is immaterial, every solution of (1) being of class C'' (hence, locally bounded and measurable).

It is known [1] that the assumption that the case $\lambda = 0$ of

(4) $$(px')' + (q + \lambda)x = 0$$

has some solution violating (3) implies that some solution of (4) must violate (3) in the case of an arbitrary $\lambda = \text{const}$. Hence, what is required is that an homogeneous, linear boundary condition assigned at the lower end of the s-half-line, *viz.*,

(5) $$x(0)\cos\phi + x'(0)\sin\phi = 0,$$

(for instance $x(0) = 0$, where $\phi = 0$) should *actually define an eigenvalue problem.*

---

* Received October 28, 1947.

[1] This is precisely the circumstance on which Weyl's alternative of *Grenzkreisfall* and *Grenzpunktfall* depends; cf. [4], p. 238.

Accordingly, the suppositions made with regard to p, q can be expressed as the following set, (*), of assumptions:

(*) ASSUMPTIONS. *Both coefficient functions p, q of (4) are continuous for $0 \leqq s < \infty$, satisfy (2), and have the property that the differential equation (4) and a boundary condition (5) determine an eigenvalue problem (with a spectrum $S = S(\phi)$ which can be, or can contain, a continuous spectrum).*

Since the expression on the left of (1) is the Lagrangian derivative of

$$L(x, x'; s) = \tfrac{1}{2}(px'^2 - qx^2),$$

the eigenvalue problem of (4) can (even though just formally) be thought of as associated with the isoperimetric problem which, with reference to the fixed boundary condition (5), depends on

$$\delta \int_0^\infty (p\xi'^2 - (q + \lambda)\xi^2)\,ds = 0.$$

Then the chances for being in the spectrum $S = S(\phi)$ appear to be "loaded" in favor of *large positive λ-values.* For, on the one hand, $p(s)$ is positive and, on the other hand, the contribution of $\xi'^2(s)$ to the last integral can be brought arbitrarily close to $+\infty$; even when $\xi = \xi(s)$ is subject to restrictions such as

$$1 = \int_0^\infty \xi^2(s)\,ds = \int_0^r \xi^2(s)\,ds \qquad (\text{so that } \xi(s) \equiv 0 \text{ for } r \leqq s < \infty),$$

which indeed fail to prevent the curve $\xi = \xi(s)$ from being arbitrarily steep. Correspondingly, and first of all because continuous spectra and cluster points of point spectra are not excluded (in this connection, cf. [6]), we were unable to use the formal connection with the heuristic minimum problem, in proving the following

THEOREM. *Under the assumptions specified under (*), the spectrum $S(\phi)$ of (4) and (5), where ϕ is arbitrarily fixed, contains values λ clustering at $\lambda = +\infty$.*

That such λ-values must cluster *either at* $+\infty$ *or at* $-\infty$ (possibly at both) is known, but this alternative lies quite on the surface. In fact, Weyl ([3], p. 47 or [4], p. 257) has pointed out that the alternative holds for *any* bounded integral kernel $K(s, t) = K(t, s)$, rather than for just those

having the particular structure of a Green function. In other words, the alternative is so general as to have nothing to do with differential equations as such. Correspondingly, even a completely continuous K can fail to have a positive characteristic value. Thus the content of the theorem to be proved is not an issue which can be settled just in Hilbert's space. Actually, the full force of Weyl's generalization of Hilb's theory for singular boundary value problems will be involved, *via* the following

LEMMA. *If, besides the assumptions specified under* (*), *it is supposed that* $S(\phi)$ *is not the entire* λ*-line, and if* λ_0 *denotes a* λ *not contained in* $S(\phi)$,

then the corresponding homogeneous differential equation

$$(px')' + (q + \lambda_0)x = 0 \tag{6}$$

possesses two real-valued, linearly independent solutions, say

$$x = u = u(s) \text{ and } x = v = v(s),$$

having a Wronskian with a signature determined by

$$(v'u - u'v)p = +1 \tag{7}$$

and assigning for $0 \leqq s < \infty$, $0 \leqq t < \infty$ *the kernel*

$$K(s, t) = u(t)v(s) = K(t, s), \text{ where } t \leqq s, \tag{8}$$

in such a way that,

for every continuous function $g = g(s)$ *of class* (L^2) *on the half-line* $0 \leqq s < \infty$, *the integral*

$$y(s) = \int_0^\infty K(s, t) g(t) dt \tag{9}$$

defines on the half-line a function which, on the one hand, is of class (L^2) *and, on the other hand, represents that solution of the inhomogeneous differential equation*

$$(py')' + (q + \lambda_0)y = g \tag{10}$$

which satisfies the given boundary condition (that to which $S(\phi)$ *belongs, i. e.,*

$$y(0) \cos\phi + y'(0) \sin\phi = 0; \tag{11}$$

cf. (5)); *finally,*

$v(s)$ is of class (L^2), i. e.,

$$\int^{\infty} v^2(s)\,ds < \infty.$$

The Lemma results by collecting and using various aspects of the general eigenvalue problem (*) which are contained in, or between the lines of, Weyl's papers [3], [4]. As this will be quite a tedious task, it will be convenient to first deduce the Theorem from the Lemma and to verify the Lemma afterwards.

Proof of the Theorem.

Suppose that the Theorem is false. Then there exists a pair (4), (5) to which there belongs an half-line $c \leqq \lambda < \infty$ containing no points of the spectrum $S(\phi)$. In particular, the Lemma is applicable at a $\lambda_0 > c$ defining this half-line. It will be shown that this leads to a contradiction.

Since the spectrum is shifted by the amount $-\mu$ if the constant μ is added to $q(s)$ in (4), it can be assumed (without altering the given boundary condition (5)) that the hypothetical c is *negative*. Hence, if $K(s,t) = K(t,s)$ is a bounded kernel defining an integral equation which has the same spectrum as the eigenvalue problem determined by (4) and (5), then the quadratic form

$$\int_0^{\infty}\int_0^{\infty} K(s,t)\xi(s)\xi(t)\,ds\,dt \tag{12}$$

cannot attain a *positive* value [2] for any continuous $\xi(s)$ which vanishes for large s. On the other hand, the kernel (8) supplied by the Lemma, a kernel which the assertions of the Lemma imply to be bounded, clearly is a Green function [3] associated with the eigenvalue problem of (4) and (5), and so the integral equation belonging to (8) determines the same spectrum as the problem of (4) and (5). Hence, if it is shown that (when $\xi(s)$ is restricted

[2] For the case of finite symmetric matrices, or of symmetric integral kernels of completely continuous type, the fact referred to is a corollary of the circumstance that the extreme values contained in the spectrum are identical with the least upper and greatest lower bounds of the values attained by the quadratic form on the unit sphere. For the case of symmetric kernels which are just bounded (in Hilbert's sense), the case occurring above, we cannot locate in the literature an explicit passage. But it is clear that the proof is exactly the same for bounded integral kernels as the proof given in [5], § 65, for bounded matrices.

[3] The verification of this fact is precisely the same as in the case of a regular Sturm-Liouville problem; cf. [2], pp. 103-107.

to continuous functions vanishing for large s) the quadratic form (12) is capable of *positive* values, it will follow that the denial of the truth of the Theorem leads to a contradiction.

In order to construct a $\xi(s)$ having the desired properties, substitute (8) into (12). Then, if $\xi(s)$ vanishes at every s not contained in an interval $a \leqq s \leqq b$, where $0 \leqq a < b < \infty$, an application of Fubini's theorem (to a continuous function on a square) reduces (12) to

$$2 \int_a^b v(s)\xi(s) \int_a^b u(t)\xi(t)\,dt\,ds. \tag{13}$$

But it turns out that the interval $a < s < b$ can be so chosen that $u(s)$ and $v(s)$ have the same sign on it. Let this be granted for a moment. It is then clear that the form (13) attains a positive value for every continuous function $\xi(s)$ which is positive for $a < s < b$ (and so, in particular, for certain continuous functions which vanish at $s = a$ and $s = b$). In fact, the Wronskian identity (7) implies that neither $u(s)$ nor $v(s)$ can vanish identically on any s-interval.

Accordingly, the proof of the Theorem will be complete if it is shown that $u(s)$ and $v(s)$ are of the same sign at any s contained in some interval, $a \leqq s \leqq b$. To this end, two cases will have to be distinguished, according as $v(s)$ does or does not have at least one zero on the open half-line, $0 < s < \infty$.

Since $x = u(s)$ and $x = v(s)$ are two linearly independent solutions of (4), they cannot have a common zero, and either of them must change its sign at any of its zeros. Consequently, if $v(s)$ is in the first case, *i. e.*, if $v(s)$ vanishes at some $s = s_0 > 0$, then $v(s)$ must be of the same sign as $u(s)$ on a sufficiently small interval either to the right or to the left of s_0. Accordingly, it is sufficient to consider the second case.

In the latter case, since $\lambda_0 > c$ is arbitrary, it follows from the continuity of $q(s)$ and from the Sturm comparison theorem that, if λ_0 is sufficiently large, every solution $x = x(s)$ of (6) vanishes at some positive s. In particular, $x = v(s)$ vanishes at some $s = s_0 > 0$. Consequently, the second case does not occur if λ_0 is suitably chosen. This completes the proof of the Theorem.

Remark. It is clear that the above proof supplies the following statement as a by-product:

If the asumptions (*) *are satisfied and if, for every* λ, *every solution of* (4) *possesses an infinity of zeros, then the spectrum* $S(\phi)$ *determined by* (4)

and (5), *where* ϕ *is arbitrarily fixed, contains values* λ *clustering at* $\lambda = -\infty$ *and at* $\lambda = +\infty$.

Proof of the Lemma.

Let the ϕ defining the boundary condition, (5) or (11), be fixed.

Since not every solution of (6) is of class (L^2) and since λ_0 is not in the spectrum $S(\phi)$, Weyl's theory supplies [4] the existence of a *real-valued* kernel

$$G(s,t) = G(t,s), \qquad (0 \leqq s < \infty, 0 \leqq t < \infty)$$

which is bounded (in the sense of the $L^2(0,\infty)$-realization of Hilbert's theory) and which has the property that

$$z_g(s) = \int_0^\infty G(s,t)g(t)\,dt \tag{14}$$

transforms every continuous $g(s)$ of class (L^2) into a function $y = z_g(s)$ of class (L^2) satisfying (10) and (11).

Needless to say, (10) cannot have (for any g) two distinct solution of class (L^2) satisfying (11). For if $y = y_1$ and $y = y_2$ are two such solutions, then $x = y_1 - y_2 \not\equiv 0$ satisfies (6), (5) and (3), which means that λ_0 is in the point spectrum, in contradiction to the assumption that λ_0 is not in the spectrum at all. The completion of the proof of the Lemma will depend on this uniqueness.

First, since λ_0 is not in $S(\phi)$, it follows from a general oscillation theorem, proved in [1], that (6) has a non-trivial solution $x(s)$ of class (L^2). Let $v = v(s)$ be such a solution (this takes care of the last assertion of the Lemma). Since λ_0, being outside the spectrum, cannot be in the point spectrum, (5) is not satisfied by $x = v(s)$. Hence, if $x = u(s)$ is any non-trivial solution of (6) and (5), then $u(s)$ is linearly independent of $v(s)$. Since nothing is altered if such a $u(s)$ is multiplied by a non-vanishing

[4] In [3], the coefficient q of the above equation (4) is supposed to be bounded from above, which, due to a theorem of A. Kneser, implies that every negative number λ of sufficiently large absolute value is "non-singular," in the sense of being a λ_0 admissible in the above Lemma (if ϕ is fixed); cf. [3], p. 39 and p. 43. In contrast, [4] does not assume that there exists such a *real* λ_0 and must, therefore, operate with a complex λ_0 $(= i)$; cf. [4], p. 267. The content of the above Lemma is the inclusion of the situation in which there exists a real λ_0 which, however, need not be supplied, as in [3], by a "Kneser half-line."

constant, and since the Wronskian, $(x_1x_2' - x_2x_1')p$, of two linearly independent solutions, $x = x_1$ and $x = x_2$, of (6) is a non-vanishing constant, it can be assumed that (7) is satisfied.

In terms of the resulting pair of solutions $v(s)$, $u(s)$ of (6), define $K(s, t)$ by (8). Then the transformation (9) of a function $g(s)$ into a $y = y_g(s)$ becomes

$$(15) \qquad y_g(s) = v(s) \int_0^s u(t)g(t)\,dt + u(s) \int_s^\infty v(t)g(t)\,dt.$$

If s, where $0 \leqq s < \infty$, is fixed, the integrals in (15) exist whenever g is of class (L^2). For, on the one hand, $u(t)$ is bounded on every finite interval $0 < t < s$ and, on the other, the product vg is of class (L), since v and g are of class (L^2).

The proof of the Lemma will be complete if it is shown that, for any continuous g of class (L^2), the functions z_g, y_g, which are defined by (14), (15), respectively, are identical. In virtue of the continuity of the function $y_g(s)$ (where g needs to be of class (L^2) only) and of the function $z_g(s)$ (where g, besides being of class (L^2), is continuous), it is sufficient to prove that, if g is any function of class (L^2), then $y_g \equiv z_g$ almost everywhere.

Clearly, (15) is nothing but the standard quadrature formula, supplying a solution of (10) satisfying (11) when two solutions of (6) are known and g is continuous. Consequently, by the uniqueness mentioned above, the identification of y_g and z_g (for a continuous g) will be complete if it is shown that y_g is of class (L^2).

To this end, suppose first that $g(s)$ is continuous and vanishes from a certain s onward, say for $s_0 \leqq s < \infty$. Then g is of class (L^2), and (15) shows that $y_g(s)$ is a constant multiple of $v(s)$ for $s_0 \leqq s < \infty$. Since v is of class (L^2), it follows that y_g is of class (L^2). Hence, $y_g(s) \equiv z_g(s)$ for such a g.

Next, the identity of y_g with z_g (almost everywhere), for an arbitrary g of class (L^2), follows from the continuity of the transformation $g \rightarrow z_g$, continuity being meant with reference to the topology defined by the $L^2(0, \infty)$-metric. For, if g is any function of class (L^2), choose a sequence of continuous functions $g_1, g_2, \cdots$ the n-th of which vanishes for $n \leqq s < \infty$ and tends to g, as $n \rightarrow \infty$, in the mean of the $L^2(0, \infty)$-space. Let z_n and y_n, respectively, denote the case $h = g_n$ of z_h and y_h. Then, as just proved, $z_n \equiv y_n$. Since the kernel $G(s, t)$ of (14) is bounded, z_n tends to z_g in the mean of (L^2). In particular, $z_1(s), z_2(s), \cdots$ contains a subsequence converging to $z_g(s)$

almost everywhere. In order to simplify the notations, n will be written in place of k_n if k_n is the n-th index of the selected subsequence. Then

$$\lim_{n\to\infty} y_n(s) = z_g(s) \text{ almost everywhere,}$$

since $y_n(s) \equiv z_n(s)$.

If g is replaced in (15) by g_n and if (15_n) denotes the resulting relation, then, as $n \to \infty$, term-by-term integration is legitimate on the right of (15_n), when s is arbitrarily fixed. For, on the one hand, g_n tends to g in the mean of $L^2(0, \infty)$, and therefore in the mean of both $L^2(0, s)$ and $L^2(s, \infty)$ and, on the other hand, v is of $L^2(0, \infty)$. Accordingly, the expression on the right of (15_n) tends to the expression on the right of (15). Since the functions on the left of (15_n) and (15) are $y_n(s)$ and $y_g(s)$, respectively, it follows that

$$\lim_{n\to\infty} y_n(s) = y_g(s)$$

holds at every s.

Since the last two formula lines imply that $y_g(s) = z_g(s)$ holds almost everywhere, the proof of the Lemma is now complete.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[2] A. Kneser, "*Die Integralgleichungen und ihre Anwendungen in der mathematischen Physik*," Braunschweig, 1922 (2nd ed.).

[3] H. Weyl, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen," *Göttinger Nachrichten*, 1909, pp. 37-63.

[4] ———, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[5] A. Wintner, "*Spektraltheorie der unendlichen Matrizen*," Leipzig, 1929.

[6] ———, "On Dirac's theory of continuous spectra," *Physical Review*, vol. 73 (1948).

ISOMORPHISMS OF JORDAN RINGS.*

By N. JACOBSON.

In a recent paper G. Ancochea [1] has defined a semi-isomorphism S of an associative ring $\mathfrak{A}$ on an associative ring $\mathfrak{B}$ to be a $1-1$ mapping of $\mathfrak{A}$ on $\mathfrak{B}$ such that

$$(1) \qquad (a+b)S = aS + bS$$

$$(2) \qquad (ab+ba)S = (aS)(bS) + (bS)(aS).$$

Moreover, Ancochea showed that if $\mathfrak{A}$ and $\mathfrak{B}$ are simple algebras with finite bases over their centers of characteristic $\neq 2$ then any semi-isomorphism of $\mathfrak{A}$ on $\mathfrak{B}$ is either an isomorphism or an anti-isomorphism. The starting point of the present paper is the observation that semi-isomorphisms are nothing more nor less than ordinary isomorphisms of the non-associative Jordan ring determined by the given associative ring. For the sake of simplicity we assume that our rings have the property that the mapping $a \rightarrow 2a$ is $1-1$. Then for each a, $\frac{1}{2}a$ is uniquely defined. Hence we can introduce the Jordan multiplication

$$a \cdot b = \tfrac{1}{2}(ab + ba)$$

If we use this in place of the ordinary multiplication we obtain the Jordan ring $\mathfrak{A}_j$ determined by the associative ring $\mathfrak{A}$. It is immediate that S is a semi-isomorphism of $\mathfrak{A}$ on $\mathfrak{B}$ if and only if this mapping is an isomorphism of $\mathfrak{A}_j$ on $\mathfrak{B}_j$.

If $\mathfrak{A}$ is a finite simple algebra it is known that $\mathfrak{A}_j$ is a simple Jordan ring. Thus from the non-associative point of view Ancochea's theorem gives a solution for one case of the general problem of determining the isomorphisms between simple Jordan rings that are algebras with a finite basis over a field. Besides the rings $\mathfrak{A}_j$ we note also the following type of simple Jordan ring. As before let $\mathfrak{A}$ be a finite simple associative algebra. Assume, moreover, that $\mathfrak{A}$ possesses an involution J, i. e. an anti-automorphism of period two, and let $\mathfrak{H}(\mathfrak{A}, J)$ be the set of elements that are J-symmetric ($aJ = a$). Then it can be shown that $\mathfrak{H}(\mathfrak{A}, J)$ is a simple (Jordan) subring of $\mathfrak{A}_j$. In this paper we determine the isomorphisms between any two simple Jordan rings of the types enumerated here ($\mathfrak{A}_j$ or $\mathfrak{H}$). As will be shown

* Received May 2, 1947.

[1] [2] in the Bibliography.

elsewhere [2] these systems together with one other type obtained from Clifford number systems give all of the simple Jordan rings that are algebras with a finite basis over a field of characteristic 0. Since it is trivial to determine the isomorphisms for Clifford systems our results give a complete solution of the isomorphism problem for simple Jordan algebras of characteristic 0.

It should be observed that the results and methods of the present paper are quite similar to our former ones on Lie rings.[3] The Jordan theory is, in fact, simpler in one important respect, namely, simple Jordan algebras have identities. Hence in studying the isomorphisms one can get along without the concept of the multiplication centralizer (extended centrum). It suffices to consider the ordinary centers for these rings. We remark finally that the results in the Lie case could also have been formulated in terms of another kind of "semi-isomorphism" of an associative ring, namely, a $1-1$ mapping S such that

$$(a+b)S = aS + bS \tag{3}$$

$$(ab)S - (ba)S = (aS)(bS) - (bS)(aS). \tag{4}$$

1. Types of Jordan rings. We shall consider the following classes of Jordan rings.

Type A: Let $\mathfrak{A}$ be a simple associative ring that has finite dimensionality over its center Φ. We assume throughout that the characteristic is not two. Let $\mathfrak{A}_j$ be the Jordan ring obtained from $\mathfrak{A}$ by replacing ordinary multiplication by the Jordan multiplication $a \cdot b = \frac{1}{2}(ab + ba)$. We shall say that $\mathfrak{A}_j$ is a Jordan ring of *type* A_I.

Assume again that $\mathfrak{A}$ is simple and finite over its center, which we now denote as P. Assume further that $\mathfrak{A}$ possesses an involution J of second kind. Thus J is an anti-automorphism of period 2 and J induces a non-trivial automorphism in P. Let $\mathfrak{H}(\mathfrak{A}, J)$ be the subring of $\mathfrak{A}_j$ of J-symmetric elements and let $\Phi = \mathfrak{H} \cap P$. Then it is known that P is a quadratic extension of Φ. The rings $\mathfrak{H}(\mathfrak{A}, J)$, J of second kind, will be called Jordan rings of *type* A_{II}.

Types B-C. The distinction between these types will be given later. Both are obtained by starting with a simple associative ring finite over the center Φ that possesses an involution J of first kind. This means that J acts

[2] This will be proved in a forthcoming joint paper on representation theory of Jordan algebras by the author and F. D. Jacobson.

[3] [3] and [4].

as the identity in Φ. Let $\mathfrak{H}(\mathfrak{A}, J)$ be the totality of J-symmetric elements. Then $\mathfrak{H}$ is a subring of $\mathfrak{A}_j$ that we shall call a Jordan ring of *type B* or *C*.

2. The extended algebras. Simplicity. If Φ is the field specified in the various cases above then the given Jordan ring $\mathfrak{U}$ contains Φ. If $\alpha \varepsilon \Phi$ the ordinary products αa and $a\alpha$ are equal. Hence

$$\alpha \cdot a = \alpha a = a\alpha. \tag{5}$$

The associative ring $\mathfrak{A}$ from which $\mathfrak{U}$ is obtained can be regarded as an algebra over Φ since

$$\alpha(ab) = (\alpha a)b = a(\alpha b) \tag{6}$$

for all $a, b \varepsilon \mathfrak{A}$. This condition also implies that

$$\alpha(a \cdot b) = (\alpha a) \cdot b = a \cdot (\alpha b) \tag{7}$$

for all $a, b \varepsilon \mathfrak{A}$. Hence $\mathfrak{U}$ can be regarded as a Jordan algebra over Φ. We now let Ω be the algebraic closure of the field Φ and we consider the extended algebras $\mathfrak{U}_\Omega$.

Type A_I. Since Ω is algebraically closed $\mathfrak{A}_\Omega = \Omega_n$ the full matrix algebra over Ω. Since $\mathfrak{U} = \mathfrak{A}_j, \mathfrak{U}_\Omega = \mathfrak{A}_{j\Omega} = (\mathfrak{A}_\Omega)_j = \Omega_{nj}$.

Type A_{II}. Here $\mathfrak{U} = \mathfrak{H}(\mathfrak{A}, J)$, J an involution of second kind. It is known that if $x_1, x_2, \cdots, x_m$ is a basis for $\mathfrak{H}$ over Φ then it is also a basis for $\mathfrak{A}$ over P. Hence m is the dimensionality n^2 of $\mathfrak{A}$ over P. This shows also that $\mathfrak{U}_P = \mathfrak{A}_j$. Hence if Ω is the algebraic closure of Φ chosen to contain P then $\mathfrak{U}_\Omega = \Omega_{nj}$.

Types B and C. In this case $\mathfrak{U} = \mathfrak{H}(\mathfrak{A}, J)$, J of first kind. The involution J can be extended to an involution $\bar{J}$ in $\mathfrak{A}_\Omega = \Omega_n$.[4] As is known we can suppose that $\bar{J}$ is either the usual mapping $a \rightarrow a'$ the transposed or $\bar{J}$ is the mapping $a \rightarrow q^{-1}a'q$ where

$$q = \begin{pmatrix} 0 & 1_\nu \\ -1_\nu & 0 \end{pmatrix}. \tag{8}$$

Of course the latter can hold only if $n = 2\nu$ is even. In the first case we say that J and $\mathfrak{H}(\mathfrak{A}, J)$ are of *type* B and in the second that these are of *type* C. If J is of type B the set $\mathfrak{H}(\Omega_n, \bar{J})$ of $\bar{J}$-symmetric elements is just the set of ordinary symmetric matrices. The dimensionality over Ω is $n(n+1)/2$. If J is of type C the set $\mathfrak{H}(\Omega_n, \bar{J})$ is the set of matrices that satisfy

[4] Cf. [3] pp. 535-537.

$$q^{-1}a'q = a. \tag{9}$$

These are of the form

$$a = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{10}$$

where the $a_{ij} \varepsilon \Omega_\nu$ and satisfy

$$a_{22} = a'_{11} \qquad a'_{12} = -a_{12}, \qquad a'_{21} = -a_{21}. \tag{11}$$

Hence the dimensionality over Ω is $n(n-1)/2 = \nu(2\nu - 1)$. For both types it is easy to see that $\mathfrak{H}(\Omega_n, \bar{J}) = \mathfrak{H}(\mathfrak{A}, J)_\Omega$.

The algebras Ω_{nj}, $\mathfrak{H}(\Omega_n, \bar{J})$, $\bar{J}$ of either type, are known to be simple.[5] It follows that the Jordan rings $\mathfrak{U}$ are simple when regarded as algebras over Φ. This implies also that *the rings* $\mathfrak{U}$ *are simple.*

3. The centers. The center Γ of any non-associative ring $\mathfrak{U}$ is defined to be the totality of elements γ that commute with every $a \varepsilon \mathfrak{U}$ and that associate with every pair a, b in the sense that

$$\gamma(a \cdot b) = (\gamma \cdot a) \cdot b, \quad a \cdot (\gamma \cdot b) = (a \cdot \gamma) \cdot b, \quad a \cdot (b \cdot \gamma) = (a \cdot b) \cdot \gamma. \tag{12}$$

It is easy to see that the middle condition is a consequence of the other two. Moreover, if the ring is commutative the first condition is sufficient that a be in the center. If we now consider a Jordan ring $\mathfrak{U}$ and we use the definition $a \cdot b = \frac{1}{2}(ab + ba)$ then we can verify that

$$[[a\gamma]b] = 0 \tag{13}$$

is equivalent to the condition $\gamma \cdot (a \cdot b) = (\gamma \cdot a) \cdot b$.

It is clear from this condition that if γ is in the field Φ specified in the various cases above then $\gamma \varepsilon \Gamma$. Hence $\Phi \subseteq \Gamma$. We shall now show that $\Gamma = \Phi$. The proof will not make use of the explicit form of (13) but rather it will depend on some general principles that are applicable to other non-associative simple rings. We observe first that because of the simplicity of $\mathfrak{U}$ the center Γ is an ordinary field.[6] The field Γ as well as $\mathfrak{U}$ is finite dimensional over Φ. Now let Ω be the algebraic closure of Φ and form the extension algebra $\mathfrak{U}_\Omega$ of $\mathfrak{U}$ regarded as an algebra over Φ. Then it is clear that $\mathfrak{U}_\Omega$ contains Γ_Ω and that the latter is in the center of $\mathfrak{U}_\Omega$. Since $\mathfrak{U}_\Omega$ is simple its center is a field. On the other hand if $(\Gamma : \Phi) > 1$ then Γ_Ω has zero-divisors. Hence we must have $(\Gamma : \Phi) = 1$ and $\Gamma = \Phi$.

---

[5] [7] pp. 7-10 or [1] p. 553.

[6] [6] p. 239 or [9] p. 62.

We can now state the results on the dimensionalities obtained in **2** in the following way. The dimensionality $(\mathfrak{U} : \Phi)$ of $\mathfrak{U}$ over its center Φ is given by the table:

$(\mathfrak{U} : \Phi) = n^2$ if $\mathfrak{U} = \mathfrak{A}_j$ and $(\mathfrak{A} : \Phi) = n^2$

$(\mathfrak{U} : \Phi) = n^2$ if $\mathfrak{U} = \mathfrak{H}(\mathfrak{A}, J)$, J of second kind and $(\mathfrak{A} : \Phi) = 2n^2$

$(\mathfrak{U} : \Phi) = n(n+1)/2$ if $\mathfrak{U} = \mathfrak{H}(\mathfrak{A}, J)$, J of type B and $(\mathfrak{A} : \Phi) = n^2$

$(\mathfrak{U} : \Phi) = n(n-1)/2$ if $\mathfrak{U} = \mathfrak{H}(\mathfrak{A}, J)$, J of type C and $(\mathfrak{A} : \Phi) = n^2$.

We shall assume in the sequel that $n > 1$ for the types A_I and B and that $n > 2$ for type C rings.

4. The enveloping rings. By the *enveloping ring* of a subset $\mathfrak{U}$ of a ring $\mathfrak{A}$ we mean the subring of $\mathfrak{A}$ generated by $\mathfrak{U}$. We shall now show that in all of the cases considered above the enveloping ring of the Jordan ring $\mathfrak{U}$ is the ring $\mathfrak{A}$ from which $\mathfrak{U}$ is constructed. This is, of course, clear if $\mathfrak{U} = \mathfrak{A}_j$ is of type A_I. Suppose next that $\mathfrak{U} = \mathfrak{H}(\mathfrak{A}, J)$ is of type A_{II} and let $\mathfrak{B}$ denote the enveloping ring. Since $P\mathfrak{H}$, the totality of linear combinations of elements of $\mathfrak{H}$ with coefficients in P, is $\mathfrak{A}$ we have also that $P\mathfrak{B} = \mathfrak{A}$. Hence if $\mathfrak{B} \supseteq P$ then $\mathfrak{B} = \mathfrak{A}$. Evidently $\mathfrak{B} \supseteq \Phi$. We assert that $\mathfrak{B}$ is a simple algebra over Φ. For $\mathfrak{B}$ is semi-simple since a nilpotent ideal $\mathfrak{N}$ in $\mathfrak{B}$ determines a nilpotent ideal $P\mathfrak{N}$ in $\mathfrak{A}$. Hence if $\mathfrak{B}$ is not simple, $\mathfrak{B} = \mathfrak{B}_1 + \mathfrak{B}_2$ where the $\mathfrak{B}_i$ are two-sided ideals. Then $\mathfrak{B}_1\mathfrak{B}_2 = 0 = \mathfrak{B}_2\mathfrak{B}_1$. If $\mathfrak{A}_i = P\mathfrak{B}_i$ then also $\mathfrak{A} = \mathfrak{A}_1 + \mathfrak{A}_2$ and $\mathfrak{A}_1\mathfrak{A}_2 = 0 = \mathfrak{A}_2\mathfrak{A}_1$. This contradicts the simplicity of $\mathfrak{A}$. Thus $\mathfrak{B}$ is simple. Now suppose that $\mathfrak{B}$ is a proper subring of $\mathfrak{A}$. Then $\mathfrak{B} \cap P = \Phi$ and it follows that Φ is the center of $\mathfrak{B}$. On the other hand J induces an anti-automorphism of period 2 in $\mathfrak{B}$. Since J leaves the elements of Φ invariant J is of first kind in $\mathfrak{B}$. Hence $\mathfrak{H} = \mathfrak{H}(\mathfrak{B}, J)$ is of type B or C. But then the dimensionality of $\mathfrak{H}$ over its center woold be $n(n+1)/2$ or $n(n-1)/2$. Since we know it is n^2 and $n > 1$ this is impossible.

Next let $\mathfrak{B}$ be the enveloping ring of $\mathfrak{U} = \mathfrak{H}(\mathfrak{A}, J)$ of type B or C. Since $\mathfrak{U}$ contains Φ, $\mathfrak{B}$ is a Φ-subalgebra of $\mathfrak{A}$. It is easy to see that the enveloping ring of $\mathfrak{U}_\Omega$ is $\mathfrak{B}_\Omega$. Also $\mathfrak{B}_\Omega = \Omega_n$ implies that $\mathfrak{B} = \mathfrak{A}$. Hence it suffices to show that the enveloping ring of $\mathfrak{U}_\Omega = \mathfrak{H}(\Omega_n, J)$ is Ω_n.

To prove this suppose first that $\bar{J}$ is of type B. Then $\mathfrak{H}(\Omega_n, \bar{J})$ has the basis $h_{ij} = (e_{ij} + e_{ji})$, $i < j$, and e_{ii} if the e_{ij} are matrix units. Then $e_{ii} \varepsilon \mathfrak{B}_\Omega$ and also e_{ij}; $e_{ii}h_{ij}e_{jj}$ and $e_{ji} = e_{jj}h_{ij}e_{ii} \varepsilon \mathfrak{B}_\Omega$. Since $\mathfrak{B}_\Omega$ contains Ω this implies that $\mathfrak{B}_\Omega = \Omega_n$. Next let $\bar{J}$ be of type C. Here $\mathfrak{H}(\Omega_n, \bar{J})$ has the basis

$$f_{ij} = e_{ij} + e_{j+\nu,i+\nu}, \qquad i, j = 1, 2, \cdots, \nu$$
$$k_{i,j+\nu} = e_{i,j+\nu} - e_{j,i+\nu}, \qquad i, j = 1, 2, \cdots, \nu, \qquad i < j$$
$$k_{i+\nu,j} = e_{i+\nu,j} - e_{j+\nu,i}, \qquad i, j = 1, 2, \cdots, \nu, \qquad i < j.$$

It follows that the enveloping ring contains $e_{ij} = f_{ii}f_{ij}$, $i \neq j$, $e_{ii} = e_{ij}e_{ji}$, $e_{j+\nu,i+\nu} = f_{ij}f_{ii}$, $i \neq j$, $e_{j+\nu,j+\nu} = e_{j+\nu,i+\nu}e_{i+\nu,j+\nu}$, $e_{i,j+\nu} = e_{ii}k_{i,j+\nu}$, $i < j$, $e_{j,i+\nu} = - e_{jj}k_{i,j+\nu}$, $i < j$, $e_{i+\nu,j} = k_{i+\nu,j}e_{jj}$, $i < j$, $e_{j+\nu,i} = - k_{i+\nu,j}e_{ii}$, $i < j$, $e_{i,i+\nu} = e_{i,j+\nu}e_{j+\nu,i+\nu}$, $e_{i+\nu,i} = e_{i+\nu,j}e_{ji}$. This shows that $\mathfrak{B}_\Omega = \Omega_n$ and completes the proof.

5. Non-isomorphim of the extended rings. We consider now two Jordan rings $\mathfrak{U}_1$ and $\mathfrak{U}_2$ or types A, B or C with enveloping rings $\mathfrak{A}_1$ and $\mathfrak{A}_2$ and centers Φ_1 and Φ_2 respectively. Let S be an isomorphism of $\mathfrak{U}_1$ on $\mathfrak{U}_2$. Then S induces an isomorphism s of Φ_1 on Φ_2. Moreover, since Jordan multiplication by elements of Φ_i coincides with ordinary multiplication, s is an isomorphism between the (ordinary) fields Φ_i. Also

$$(\gamma a_1)S = (\gamma \cdot a_1)S = \gamma^s \cdot a_1 S = \gamma^s(a_1 S). \tag{14}$$

By constructing a new system $\mathfrak{A}_2$ if necessary we can suppose that $\Phi_1 = \Phi_2 = \Phi$. Then s is an automorphism in Φ and by (14) S is a semi-linear transformation between $\mathfrak{U}_1$ and $\mathfrak{U}_2$ regarded as vector spaces over Φ.

As before let Ω denote the algebraic closure of Φ. Then s can be extended to an automorphism $\bar{s}$ in Ω and S can be extended in one and only one way to a semi-linear transformation $\bar{S}$ of $\mathfrak{U}_{1\Omega}$ on $\mathfrak{U}_{2\Omega}$ having $\bar{s}$ as associated automorphism. It is easy to see that $\bar{S}$ is an isomorphism between the Jordan rings $\mathfrak{U}_{i\,\Omega}$.

We have seen that if $\mathfrak{U}$ is one of our Jordan rings then $\mathfrak{U}_\Omega$ is either of the form Ω_{nj} or of the form $\mathfrak{H}(\Omega_n, \bar{J})$ where J is either of type B or of type C. Now it has been proved by Albert[7] that distinct systems of the forms noted are not isomorphic. We give a new proof of this fact here.

We shall define the *degree* of a simple Jordan ring over its center to be the maximum dimensionality of any subspace A generated by the powers of an element a in the ring. It is clear from the definition of Jordan multiplication that Jordan powers coincide with ordinary powers. Also we know that Ω is the center of Ω_{nj} and of $\mathfrak{H}(\Omega_n, \bar{J})$ and that Jordan multiplication by elements of Ω is the same as ordinary multiplication. Hence the degree of Ω_{nj} and of $\mathfrak{H}(\Omega_n, \bar{J})$ is the maximum degree of the minimum polynomials of the elements of the system.

[7] [1] p. 554.

By the Hamilton-Cayley theorem it is clear that the degree $\deg \Omega_{nj} \leq n$ and $\deg \mathfrak{H}(\Omega_n, \tilde{J}) \leq n$. It is known also that if a is a $2\nu \times 2\nu$ matrix such that $a = q^{-1}a'q$ where q is skew-symmetric then a satisfies an equation of degree ν.[8] Hence $\deg \mathfrak{H}(\Omega_n, \tilde{J}) \leq \nu$ if $n = 2\nu$ and J is of type C. Since Ω is infinite, Ω_n contains a matrix

$$\operatorname{diag}\{\rho_1, \rho_2, \cdots, \rho_n\}$$

with $\rho_i \neq \rho_j$ for $i \neq j$ and $\mathfrak{H}(\Omega_n, \tilde{J})$ of type C contains

$$\operatorname{diag}\{\rho_1, \rho_2, \cdots, \rho_\nu; \rho_1, \rho_2, \cdots, \rho_\nu\}.$$

Hence the bounds indicated for the degrees are attained:

$$\deg \Omega_{nj} = n, \ \deg \mathfrak{H}(\Omega_n, \tilde{J}) = n \text{ or } \nu = n/2$$

according as $\tilde{J}$ is of type B or of type C.

Now suppose that two of these Jordan rings are isomorphic. Then the degrees are the same. Hence the isomorphic pair is in the triple: Ω_{nj}, $\mathfrak{H}(\Omega_n, \tilde{J})$, $\tilde{J}$ of type B; $\mathfrak{H}(\Omega_{2n}, \tilde{J})$, $\tilde{J}$ of type C. The dimensionalities over the center are, respectively, n^2, $n(n+1)/2$, $n(2n-1)$. Since we have assumed $n > 1$ and $2n > 2$ respectively in the last two cases, these numbers are all different. Hence the isomorphic pair coincide.

Our result shows that if $\mathfrak{U}_1$ and $\mathfrak{U}_2$ are isomorphic Jordan rings of types A, B or C then they have the same type and we can suppose that $\mathfrak{U}_{1\Omega} = \mathfrak{U}_{2\Omega}$ for Ω the algebraic closure of the center Φ. In this case, of course, the extended isomorphism $\bar{S}$ is an automorphism.

6. Isomorphisms between Jordan rings of type A. We suppose now that $\mathfrak{U}_1$ and $\mathfrak{U}_2$ are of type A. We assume first that $\mathfrak{U}_1$ is of type A_I and $\mathfrak{U}_2$ is of type A_{II}. Then $\mathfrak{U}_1 = \mathfrak{A}_j$ for a simple ring $\mathfrak{A}$ with center Φ and $\mathfrak{U}_2 = \mathfrak{H}(\mathfrak{B}, J)$ where $\mathfrak{B}$ is simple with center P a quadratic extension of Φ and J is of the second kind. Now $\mathfrak{A}_\Omega = \Omega_n$. Hence we can suppose that $\mathfrak{A}$ is a subring of the associative ring Ω_n. Similarly since $(\mathfrak{B}$ over $P)_\Omega = \Omega_n$ we can suppose that $\mathfrak{B}$ is a subring of Ω_n.

As above let S be an isomorphism of $\mathfrak{U}_1$ on $\mathfrak{U}_2$ and let $\bar{S}$ be the extended automorphism in $\mathfrak{U}_{1\Omega} = \mathfrak{U}_{2\Omega} = \Omega_{nj}$. The form of $\bar{S}$ can be deduced from the following

[8] [5] p. 748 or [4] p. 497.

LEMMA 1 (Ancochea). *Any algebra automorphism of Ω_{nj} over Ω is either an automorphism or an anti-automorphism of the associative algebra Ω_n over Ω.*[9]

By an algebra automorphism T we mean a ring automorphism such that $(\omega a)S = \omega(aS)$ for all $\omega \in \Omega$. The proof is a simple computation with matrix units and will be omitted.

We consider now the automorphism $\bar{S}$ in Ω_{nj}. We have $(\omega a)\bar{S} = \omega^{\bar{s}}(aS)$. The mapping $(\alpha_{ij}) \rightarrow (\alpha_{ij}{}^{\bar{s}})$ is an automorphism G in the associative system Ω_n. Hence it is an automorphism in Ω_{ni}. Since $(\omega a)G = \omega^{\bar{s}}(aG)$, $T = \bar{S}G^{-1}$ is an algebra automorphism in Ω_{ni}. By Lemma 1, T is an automorphism or an anti-automorphism of Ω_n. Hence $\bar{S} = TG$ is either an automorphism or an anti-automorphism of Ω_n. Clearly $\bar{S}$ induces either an isomorphism S_1 or an anti-isomorphism S'_1 of the enveloping rings of $\mathfrak{U}_1$ on the enveloping ring of $\mathfrak{U}_2$. We know that these rings are respectively $\mathfrak{A}$ and $\mathfrak{B}$. Since S_1 or S'_1 coincides with S on $\mathfrak{U}_1$ and $\mathfrak{U}_1 = \mathfrak{A}$ the image $\mathfrak{U}_1 S$ is closed under multiplication. Hence $\mathfrak{U}_1 S = \mathfrak{B}$. Since $\mathfrak{U}_1 S = \mathfrak{U}_2 = \mathfrak{H}(\mathfrak{B}, J)$ this implies that the elements of $\mathfrak{B}$ are invariant under J contrary to assumption. We therefore have the following

THEOREM 1. *A simple Jordan ring of type* A_I *can not be isomorphic to one of type* A_{II}.

We suppose next that both $\mathfrak{U}_1$ and $\mathfrak{U}_2$ are of type A_I, say, $\mathfrak{U}_1 = \mathfrak{A}_j$ and $\mathfrak{U}_2 = \mathfrak{B}_j$. Also we can suppose that $\mathfrak{A}$ and $\mathfrak{B}$ are subrings of Ω_n. Then the argument just given shows that S is either an isomorphism or an anti-isomorphism of $\mathfrak{A}$ on $\mathfrak{B}$. The converse is clear: If $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic (anti-isomorphic) and S is an isomorphism (anti-isomorphism) of $\mathfrak{A}$ on $\mathfrak{B}$ then $\mathfrak{A}_j$ and $\mathfrak{B}_j$ are isomorphic under S. This proves

THEOREM 2 (Ancochea). *If $\mathfrak{A}$ and $\mathfrak{B}$ are simple associative rings that have finite dimensionalities over their centers then the Jordan rings $\mathfrak{A}_j$ and $\mathfrak{B}_j$ are isomorphic if and only if $\mathfrak{A}$ and $\mathfrak{B}$ are either isomorphic or anti-isomorphic. Any automorphism in $\mathfrak{A}_j$ is either an automorphism or an anti-automorphism of $\mathfrak{A}$.*

We assume finally that both $\mathfrak{U}_1$ and $\mathfrak{U}_2$ are of type A_{II}, say. $\mathfrak{U}_1 = \mathfrak{H}(\mathfrak{A}, J)$ and $\mathfrak{U}_2 = \mathfrak{H}(\mathfrak{B}, K)$. Here we can suppose that $\mathfrak{A}$ and $\mathfrak{B}$ are subrings of Ω_n. Our argument shows that if S is an isomorphism of $\mathfrak{U}_1$ on $\mathfrak{U}_2$ there exists either an isomorphism S_1 or an anti-isomorphism S'_1 of $\mathfrak{A}$ on $\mathfrak{B}$ that induces

[9] [2] pp. 151-152.

S in $\mathfrak{U}_1$. We note that in reality both S_1 and S'_1 exist. For if S_1 is an isomorphism of the required type then S_1K and JS_1 are anti-isomorphisms having the required properties. Similarly if S'_1 is given then S'_1K and JS'_1 are isomorphisms. Since $\mathfrak{A}$ and $\mathfrak{B}$ are the enveloping rings of $\mathfrak{U}_1$ and $\mathfrak{U}_2$ respectively there is only one isomorphism and only one anti-isomorphism that coincides with S in $\mathfrak{U}_1$. Hence we see that $S_1K = JS_1$ and $S'_1K = JS'_1$.

We summarize our results as follows: If $\mathfrak{H}(\mathfrak{A}, J)$ and $\mathfrak{H}(\mathfrak{B}, K)$ of type A_{II} are isomorphic then $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic. In this case we can identify $\mathfrak{A}$ and $\mathfrak{B}$ and consider $\mathfrak{H}(\mathfrak{A}, J)$ and $\mathfrak{H}(\mathfrak{A}, K)$. Then a necessary condition for isomorphism is that the involutions J and K be *cogredient* in the sense that $K = S_1^{-1}JS_1$ where S_1 is an automorphism in $\mathfrak{A}$. It is easy to see that this condition is also sufficient for isomorphism. Finally we see that any automorphism S of $\mathfrak{H}(\mathfrak{A}, J)$ is induced by an automorphism S_1 of $\mathfrak{A}$ that commutes with J. For in this case we have $J = S_1^{-1}JS_1$. The automorphism S_1 is uniquely determined by S. Conversely any S_1 that commutes with J induces an automorphism in $\mathfrak{H}(\mathfrak{A}, J)$.

THEOREM 3. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be simple associative rings that have finite dimensionalities over their centers and that possess involutions J and K, respectively, of second kind. Then if the Jordan rings $\mathfrak{H}(\mathfrak{A}, J)$ and $\mathfrak{H}(\mathfrak{B}, K)$ are isomorphic, $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic. A necessary and sufficient condition that $\mathfrak{H}(\mathfrak{A}, J)$ and $\mathfrak{H}(\mathfrak{A}, K)$ be isomorphic is that J and K be cogredient. The group of automorphisms of $\mathfrak{H}(\mathfrak{A}, J)$ is the subgroup of the group of automorphisms of $\mathfrak{A}$ that commute with J.*

7. Isomorphisms between Jordan rings of types B and C. The isomorphism theory for these rings is similar to that of the rings of type A_{II}. Let $\mathfrak{U}_1 = \mathfrak{H}(\mathfrak{A}, J)$ and $\mathfrak{U}_2 = \mathfrak{H}(\mathfrak{B}, K)$, J and K of first kind (types B or C) and let S be an isomorphism of $\mathfrak{U}_1$ on $\mathfrak{U}_2$. Then we can suppose that $\mathfrak{A}$ and $\mathfrak{B}$ have the same center Φ. Let the dimensionality of $\mathfrak{A}$ over Φ be n^2. Then $\mathfrak{A}_\Omega = \Omega_n$ and $\mathfrak{U}_{1\Omega} = \mathfrak{H}(\Omega_n \bar{J})$, $\bar{J}$ of type B or type C. Similarly if the dimensionality of $\mathfrak{B}$ over Φ is m^2 then $\mathfrak{B}_\Omega = \Omega_m$ and $\mathfrak{U}_{2\Omega} = \mathfrak{H}(\Omega_m, \bar{K})$. It follows that $m = n$ and that $\bar{J} = \bar{K}$. Thus we can suppose that $\mathfrak{U}_{1\Omega} = \mathfrak{U}_{2\Omega} = \mathfrak{H}(\Omega_n, \bar{J})$ and that the enveloping rings $\mathfrak{A}$ and $\mathfrak{B}$ are subrings of Ω_n.

The isomorphism S can be extended to an automorphism $\bar{S}$ of the system $\mathfrak{H}(\Omega_n, \bar{J})$. We can prove that $\bar{S}$ is induced by an automorphism of Ω_n by using the following

LEMMA 2 (Kalisch). *Any algebra automorphism of the Jordan algebra*

$\mathfrak{H}(\Omega_n, \bar{J})$ *over* Ω *can be extended to an automorphism in the associative algebra* Ω_n.[10]

As before this result can be used to show that $\bar{S}$ can be extended to an isomorphism in Ω_n. It follows that $\bar{S}$ induces an isomorphism S_1 between the enveloping rings $\mathfrak{A}$ and $\mathfrak{B}$. The rest of the argument is a duplication of the one given in the A_{II} case. The following result which has been proved for algebras by Kalisch is therefore readily established.

THEOREM 4. *Theorem 3 holds also for involutions J and K of the first kind.*

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY

1. A. A. Albert, "On Jordan algebras of linear transformations," *Transactions of the American Mathematical Society*, vol. 59 (1946), pp. 524-555.
2. G. Ancochea, "On semi-automorphisms of division algebras," *Annals of Mathematics*, vol. 48 (1947), pp. 147-154.
3. N. Jacobson, "Simple Lie algebras over a field of characteristic zero," *Duke Mathematical Journal*, vol. 4 (1938), pp. 534-551.
4. ———, "Classes of restricted Lie algebras of characteristic p," I, *American Journal of Mathematics*, vol. 63 (1941), pp. 481-515.
5. ———, "An application of E. H. Moore's determinant of a hermitian matrix," *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 745-748.
6. ———, "Structure theory of simple rings without finiteness assumptions," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 228-245.
7. G. Kalisch, "On special Jordan algebras," *Transactions of the American Mathematical Society*, vol. 61 (1947), pp. 482-494.
8. I. Kaplansky, "Semi-automorphisms of rings," *Duke Mathematical Journal*, vol. 14 (1947), pp. 521-527.
9. T. Nakayama, "Über einfache distributive Systeme unendlicher Ränge," *Proceedings of the Imperial Academy, Tokyo*, vol. 20 (1944), pp. 61-66.

---

[10] A more general result will be proved in the paper referred to in [2].

SUBGROUPS OF LOCALLY COMPACT GROUPS.*

By Deane Montgomery.

1. Introduction. Very little is known about the structure of a locally compact group unless the group is compact or abelian. It is the purpose of this paper to prove a theorem (Theorem 2) about n-dimensional groups which is suggested intuitively and which will be useful for further investigation. Theorem 1, a theorem about more general locally compact groops, is needed in the proof of Theorem 2 and is of interest in itself. After writing this paper the author learned that Leo Zippin has independently discovered Theorem 1. The two theorems are stated as follows:

Theorem 1. *Let G be a locally compact connected group and let U be an open subset of G which includes the identity e. If K is the identity component of U and if M is the smallest closed subgroup generated by K, then M is all of G.*

Theorem 2. *Let G be a locally compact connected group of dimension n. If H is an n-dimensional closed subgroup of G, then H is all of G.*

Since dimension theory is in an unsatisfactory state beyond the range of separable metric spaces, the topology of G is assumed to be separable metric. In case G is compact or abelian the two theorems follow from the work of von Neumann and Pontrjagin [5, 4] on the structure of these repective types of groups. However, in the general case these theorems do not follow from any known result.

2. Proof of Theorem 1. The proof of Theorem 1 is made by assuming that the theorem is false and showing that this assumption leads to a contradiction. It will therefore be assumed that M is not all of G, but is a proper subgroup.

The mapping

$$f\colon\ G \to G/M$$

is continuous and open. The space G/M is connected and by the assumption above must contain more than one point. It is therefore at least one-dimensional and it is also a separable metric locally compact space.

* Received July 22, 1947.

Let V be an open set in G which satisfies the following conditions:

(a) e is in V

(b) $V = V^{-1}$

(c) $\bar{V}^2 \subset U$

(d) $\bar{V}$ is compact.

Let a be any point of $\bar{V}$ and let A be the component of $aM \cap \bar{V}$ which includes a. In symbols this will be written as

$$A = C_a(aM \cap \bar{V}).$$

By definition $K = C_e(U) = C_e(M \cap U)$ and therefore $aK = C_a(a(M \cap U)) = C_a(aM \cap aU)$. But from the conditions on V it follows that $\bar{V}$ is in aU and hence $C_a(aM \cap \bar{V}) \subset C_a(aM \cap aU)$ that is $A \subset aK$, and $A = C_a(aK \cap \bar{V})$. Now let $Q = C_a(\bar{V})$ so that it is immediate that $A \subset Q$. By the choice of V, $a^{-1}Q \subset a^{-1}\bar{V} \subset U$ and since $a^{-1}Q$ includes e it follows that $a^{-1}Q \subset K$ and $Q \subset aK$. From this it can be seen that $Q \subset A$ and hence $Q = A$. In view of the definitions it has therefore been shown that

$$C_a(\bar{V}) = C_a(aM \cap \bar{V}).$$

When $b \varepsilon aM$, then $bM = aM$ and hence when $b \varepsilon aM\bar{V}$, then

$$C_b(\bar{V}) = C_b(aM \cap \bar{V}).$$

The components of $\bar{V}$ form an upper semi-continuous collection and they determine a space B and a continuous map T [6; p. 126] of $\bar{V}$ onto B

$$T: \bar{V} \rightarrow B$$

such that for every b in B the set $T^{-1}(b)$ is a component of $\bar{V}$. The space B is compact and metric and must be totally disconnected (= zero dimensional). In order to see this last point notice that since T is monotone the inverse of any connected set in B is connected. Hence, if B contained a connected set with more than one point, its inverse would be a connected set containing more than one component of $\bar{V}$ which is impossible. This proves that B is totally disconnected.

The fact that B is totally disconnected implies that if Y^* is any open set (in and relative to B) including a closed set B^*, then there is a subset X^* of Y^* which includes B^* and is both open and closed in B [1, Chap. II].

Again let a be any point of $\bar{V}$ and consider the closed set $aM \cap \bar{V}$. By previous results this set is the union of a certain set of components of $\bar{V}$, so that $T^{-1}[T(aM \cap \bar{V})] = aM \cap \bar{V}$. Let

$$T(aM \cap \bar{V}) = B^*,$$

and let Y be any open set in and relative to $\bar{V}$ such that

$$aM \cap \bar{V} \subset Y.$$

Let Y_1 be the set of all points of Y each of which is on a component of $\bar{V}$ entirely contained in Y. Then [6; p. 123] Y_1 is open in and relative to $\bar{V}$. Furthermore, the set $T(Y_1)$ is, by definition, open in B and $B^* \subset T(Y_1)$. Consequently, there is a set X^* which is such that

(a) $B^* \subset X^* \subset T(Y_1)$

(b) X^* is both open and closed in B.

The set $X = T^{-1}(X^*)$ is open and closed in $\bar{V}$ and X is such that $AM \cap \bar{V} \subset X \subset Y_1 \subset Y$.

To sum up, *it has now been shown that if Y is any open set, in and relative to $\bar{V}$, and $aM \cap \bar{V} \subset Y$, then there is a set X open and closed in $\bar{V}$ and such that $aM \cap \bar{V} \subset X \subset Y$.*

Let W be any open set of V such that $\bar{W} \subset V$, and consider the set $f(W)$ which is an open subset of G/M. The space G/M is the union of a countable number of sets each of which is homeomorphic to $f(W)$ and $f(W)$ is the union of a countable number of compact sets. If each of these compact subsets of $f(W)$ was of dimension zero, then by the sum theorem [1] $f(W)$ and G/M would be zero dimensional, but G/M is not zero dimensional. Hence $f(W)$ contains a compact set of dimension greater than zero, and $f(W)$ is of dimension greater than zero. The contradiction which is being sought will be obtained by showing that, on the other hand, $f(W)$ must be zero dimensional.

Let w^* be any point of $f(W)$ and let w in W be such that $f(w) = w^*$. Let Y^* be any set in and open in $f(W)$ such that $w^* \varepsilon Y^*$ and let $Y = f^{-1}(Y^*) \cap \bar{V}$. Then Y is in and open in $\bar{V}$ and $wM \cap \bar{V}\, Y$. By the result obtained above there is a set X open and closed in $\bar{V}$ which is such that $wM \cap \bar{V} \subset X \subset Y$. It is also true that $wM \cap W \subset X$ and hence $f(wM \cap W) \subset f(X) \subset f(W)$. The set $f(X) \cap f(W)$ is closed in $f(W)$ and it will next be proved that it is also open in $f(W)$.

Assume that b^* is any point of $f(X) \cap f(W)$. Choose any point b such that $b \varepsilon W \cap X, f(b) = b^*$. Then b is also in a set O which is such that

1) O is open in G

2) $O \subset W$

3) $O \subset X$.

Hence, $f(O)$ is open in G/M, and $f(O)$ is in and open in $f(W)$ and $f(X)$. But b^* is in $f(O)$, that is $b^* \varepsilon f(O) \subset f(X) \cap f(W)$ which proves that the arbitrary point b^* of $f(X) \cap f(w)$ is in a set open in G which is in $f(X) \cap f(W)$. This proves that $f(X) \cap f(W)$ is open in $f(W)$.

It has therefore been shown that w^*, an arbitrary point in $f(W)$, is in a subset $f(X) \cap f(W)$ of $f(W)$ which is open and closed in $f(W)$ and which is in a given open set Y^* containing w^*. This proves that $f(W)$ is zero dimensional and gives the desired contradiction. This completes the proof of Theorem 1.

A *constituent* of a point x in a topological space is defined to be the totality of all points y such that x and y are in a compact connected subset.

COROLLARY 1. *Let G be a locally compact connected group. Then the constituent of the identity is everywhere dense in G.*

This follows from the fact that $\bar{K}$, $\bar{K}^2$, $\bar{K}^{-1}$, $\bar{K}^{-2}$ and so on are compact connected sets whose union is dense in M and hence in G.

COROLLARY 2. *Let G be a locally compact connected group and let O be an open set including e. If H is a closed proper subgroup of G, then O contains a continuum C such that e is in C but C is not in H.*

In the previous proof the set $\bar{K}$ is a continuum in $\bar{U}$ which contains e If H is a proper subgroup then $\bar{K}$ is not in H. If U is chosen so that $\bar{U}$ is in O the corollary follows.

3. Non-existence of n-cycles. It is convenient to prove a lemma about the non-existence of n-cycles, which is entirely analogous to a lemma already stated in the one-dimensional case [3]. However, the proof is sketched for the sake of completeness, and is made simpler by Theorem 1.

LEMMA 1. *Let G be a locally compact connected n-dimensional group which is not compact. Then if B is any compact set in G the n-th homology group of B (reals* mod 1) *is trivial.*

If B is any compact set in G, then by Corollary 1, there must be a compact connected set Q which contains e and which contains a point q which is such that B and $q(B)$ do not intersect. Let z be any Vietoris n-cycle in B which, contrary to the desired conclusion, does not bound in B. Then $q(z)$ is a Vietoris n-cycle which does not bound in $q(B)$, and $q(z) - z$ is a Vietoris n-cycle in $B \cup q(B)$ which does not bound in $B \cup q(B)$.

However, it can be proved exactly as before [3] that $q(z) - z$ does bound in $Q(B)$. This is done by choosing a finite chain of points $e = q_1, \cdots, q_n = q$ with $d(q_i, q_{i-1}) < r$ where r is a preassigned positive number. Using these intermediate points and the standard constructions of algebraic topology, it is proved that $q(z) - z$ bounds in $Q(B)$.

The fact that $q(z) - z$ bounds in $Q(B)$ but not in $B \cup q(B)$ contradicts the hypothesis that G is n-dimensional [1, p. 151]. This contradiction completes the proof of the lemma.

4. Proof of Theorem 2. In proving Theorem 2 it will be assumed that the conclusion is false and that H is not all of G. In more detail, it is now assumed that G is a locally compact, connected, n-dimensional group and that H is a closed n-dimensional subgroup which is not all of G. It will be shown that this assumption leads to a contradiction.

The n-dimensional group H must contain a compact set F_1 which is also n-dimensional, and the set F_1 must therefore contain a closed set B, $B \subset F_1$, with the property that there is an $(n-1)$-dimensional cycle z (reals modulo one as coefficients) which bounds in F_1 but not in B [1]. Then there exists a neighborhood U of e in G which is such that z does not bound in $\bar{U}B$.

By Corollary 2 there is a compact connected set C with the following properties $C \subset U$, $e \in C$, $C \not\subset H$. Let c be a point of C which is not in H, and let

$$
\begin{aligned}
F_2 &= cF_1 \cup CB \\
F &= F_1 \cup F_2 \\
F_{12} &= F_1 \cap F_2.
\end{aligned}
$$

Then $B \subset F_{12}$ so that the cycle z is carried by F_{12}. But

(1) $$z \not\sim 0 \text{ in } F_{12}$$

because F_{12} is in UB. However, it is known that

(2) $$z \sim 0 \text{ in } F_1$$

and it is also true that

(3) $$z \sim 0 \text{ in } F_2.$$

This last fact can be seen in the following way. As in the preceding section $cz \sim z$ in CB because C is a compact connected set including e and c, and $cz \sim 0$ in CF_1 since cF_1 is homeomorphic to F_1. From these remarks (3) follows.

From homology properties of unions and intersections of sets [2, pp. 266-271, particularly (18. 5) and p. 270. The equivalence of the Čech theory of homology used in 18. 5 and the Vietoris theory is also proved in [2].] the desired contradiction can now be obtained. The relation (18. 5) of [2] is $H^n/D^n = L^{n-1}$ where H^n is the n-dimensional homology group of F, D^n is the subgroup of H^n consisting of the differences $\gamma_1{}^n - \gamma_2{}^n$ where $\gamma_i{}^n$ is a homology class in F_i, and L^{n-1} is made up of the $(n-1)$-dimensional homology classes in F_{12} which are homologous to zero in both F_1 and F_2. It has been shown above that L^{n-1} is not a trivial group. On the other hand, Lemma 1 proves that H^n and D^n are trivial, and this contradicts the relation above. This contradiction completes the proof of Theorem 2.

YALE UNIVERSITY.

BIBLIOGRAPHY

1. Hurewicz and Wallman, *Dimension Theory*, Princeton University Press, 1941.
2. Lefschetz, Algebraic Topology, *American Mathematical Society Colloquium Publication*, vol. 27, New York, 1942.
3. Montgomery, "Connected one dimensional groups," *Annals of Mathematics*, vol. 49 (1948), pp. 110-117.
4. Pontrjagin, *Topological Groups*, Princeton University Press, 1939.
5. von Neumann, "Die Einfuhrung analytischer Parameter in topologische Gruppen," *Annals of Mathematics*, vol. 34 (1933), pp. 170-190.
6. Whyburn, Analytic Topology, *American Mathematical Society Colloquium Publication*, vol. 28, New York, 1942.

DIFFERENTIAL GEOMETRY OF A SURFACE AT A PARABOLIC POINT.*

By Chuan-Chih Hsiung.[1]

1. Introduction. In recent years projective differential geometers have made interesting contributions to singular points of curves and surfaces.[2] Since at an ordinary point of a surface in ordinary space there are two distinct asymptotic tangents, a singular point which naturally presents itself for consideration is the parabolic point, at which the two asymptotic tangents are coincident. The purpose of this paper is to study the projective differential geometry of a surface at a parabolic point. We shall speak of the plane section of a surface made by its tangent plane at a parabolic point as *the tangential section.* According as the tangential section has a cusp or a tacnode at the parabolic point, there are five essentially different cases. For each case there is obtained a canonical power series expansion of the surface in the neighborhood of the point, together with a geometrical interpretation of the system of reference giving rise to the expansion.

By the use of the osculants associated with a cusp of a plane curve, Popa [10] has obtained a canonical power series expansion of a surface at a parabolic point for the case of a cuspidal tangential section. In **2** we shall derive another simple canonical expansion for this case in a different manner.

2. Tangential section with a cusp. First of all we establish a point coordinate system in ordinary projective space, in which a point has non-homogeneous coordinates x, y, z and homogeneous coordinates $x_1, \cdots, x_4$ connected by the relations

$$x = x_2/x_1, \qquad y = x_3/x_1, \qquad z = x_4/x_1.$$

Let O be a parabolic point on an analytic surface S, whose equation is

$$z = f(x, y).$$

If we choose the point O as origin, the coincident asymptotic tangents and the tangent plane of the surface S at the point O respectively as the x-axis and

* Received September 22, 1947; Presented to the Society, April 26, 1947.

[1] The author wishes to thank Professor V. G. Grove for his suggestions in the preparation of this paper.

[2] See the bibliography at the end of the paper.

the coordinate plane $z=0$, then the Taylor's series expansion of the function $f(x,y)$ in the neighborhood of the point O may be written in the form

$$(1)\qquad z = ay^2 + b_1x^3 + b_2x^2y + b_3xy^2 + b_4y^3 + c_1x^4 + c_2x^3y + c_3x^2y^2 + c_4xy^3 + c_5y^4 + \cdots \qquad (a \neq 0),$$

the unwritten terms being of degree at least five. In this section we shall consider the general case, $b_1 \neq 0$, in which the tangential section at the point O of the surface S has a cusp at O.

It is known that for a non-asymptotic tangent at an ordinary point of a surface there is a quadric of Moutard, which is the locus of the osculating conic at the point of the plane section of the surface made by a variable plane through the tangent. For a parabolic point, the quadric degenerates into a cone with vertex on the coincident asymptotic tangents [10]. By means of the expansion (1), it is easily seen that the cone of Moutard for the y-axis at the point O of the surface S has the equation

$$(2)\qquad z - ay^2 - (b_3/a)xz - (b_4/a)yz + (1/a^2)(b_4^2/a - c_5)z^2 = 0,$$

with vertex at the point $(b_3, a, 0, 0)$. If we choose this vertex as the vertex $(0,1,0,0)$ of the tetrahedron of reference, then

$$(3)\qquad b_3 = 0.$$

Furthermore, if the vertex $(0,0,1,0)$ of the tetrahedron is taken at the point of intersection of the y-axis and the polar plane of the vertex $(0,0,0,1)$ with respect to the cone (2) of Moutard, then

$$(4)\qquad b_4 = 0.$$

If the vertex $(0,0,0,1)$ and the unit point of the system of reference are on the cone (2) then, respectively,

$$(5)\qquad c_5 = 0, \qquad a = 1.$$

After these simplifications, expansion (1) and equation (2) become respectively

$$(6)\qquad z = y^2 + b_1x^3 + b_2x^2y + c_1x^4 + c_2x^3y + c_3x^2y^2 + c_4xy^3 + d_1x^5 + d_2x^4y + \cdots + d_6y^5 + \cdots,$$

$$(7)\qquad z = y^2.$$

Through the y-axis there are two plane sections of the surface S each of

which has a sextactic point or a six-point conic at the point O. From (6) these two planes are easily found to be given by the following equation

(8) $$b_2x^2 + c_4xz + d_6z^3 = 0.$$

If the coordinate plane $x = 0$ is the harmonic conjugate of the tangent plane $z = 0$ with respect to the two planes (8), then

(9) $$c_4 = 0.$$

A general plane through the asymptotic tangent $y = z = 0$ cuts the surface S in a curve having an inflexion at the point O. Associated with any such plane section, we may determine Bompiani's covariant point O_1. Its coordinates [1] are

(10) $$(c_1, 2b_1, 0, 0),$$

and hence are independent of the plane considered. If the point (10) is taken as the vertex $(0, 1, 0, 0)$, then $c_1 = 0$.

Let us now consider a general tangent of the surface at the point O; its equations are

(11) $$z = y - \lambda x = 0 \qquad (\lambda \neq 0).$$

Any plane through the tangent (11),

(12) $$z = n(y - \lambda x) \qquad (n \neq 0),$$

cuts the surface S in a curve C. The expansion of the projection C' of the curve C from the vertex $(0, 0, 0, 1)$ onto the tangent plane $z = 0$ is found, by eliminating z between equations (6), (12), to be

(13) $$y = \lambda x + (\lambda^2/n)x^2 + (1/n)(b_1 + b_2\lambda + 2\lambda^3/n)x^3 + (1/n)(5\lambda^4/n^2 + 3b_2\lambda^2/n + 2b_1\lambda/n + c_1 + c_2\lambda + c_3\lambda^2 + c_4\lambda^3)x^4 + \cdots.$$

The osculating conic of the curve C' at the point O is given by the equations $z = 0$ and

(14) $$\begin{aligned} Q \equiv y - \lambda x &- (\lambda^2/n)x^2 - (1/\lambda^2)(2\lambda^3/n + b_2\lambda + b_1)x(y - \lambda x) \\ &- (n/\lambda^4)(\lambda^4/n^2 - b_2\lambda^2/n - 2b_1\lambda/n - b_2{}^2 - b_1{}^2/\lambda^2 \\ &- 2b_1b_2/\lambda + c_1 + c_2\lambda + c_3\lambda^2 + c_4\lambda^3)(y - \lambda x)^2 = 0. \end{aligned}$$

Elimination of n between equations (12), (14) yields the equation of the cone of Moutard for the given tangent (11)

$$(15)\qquad z - y^2 + (1/\lambda^3)(2b_1 + b_2\lambda)yz - (1/\lambda^2)(3b_1 + 2b_2\lambda)xz \\ + (1/\lambda^6)[b_1^2 + 2b_1b_2\lambda + (b_2^2 - c_1)\lambda^2 - c_2\lambda^3 - c_3\lambda^4 - c_4\lambda^5]z^2 = 0.$$

The residual conic of intersection of this cone with the other cone of Moutard associated with the harmonic conjugate of the tangent (11) with respect to the x-, y-axes lies in the plane

$$(16)\qquad 2b_2\lambda^2x - 2b_1y - (1/\lambda^2)(2b_1b_2 - c_2\lambda^2)z = 0.$$

The plane through the asymptotic tangent $y = z = 0$ and the line of intersection of the two planes (16) and $x = 0$ is given by the equation

$$(17)\qquad 2b_1y + (1/\lambda^2)(2b_1b_2 - c_2\lambda^2)z = 0.$$

The equations of a general line through the point O and in the plane $x = 0$ may be written as

$$(18)\qquad x = z - ny = 0,$$

so that the plane (12) passes through the line and a variable tangent (11). As the tangent (11) varies through the point O, the locus of the line of intersection of the two planes (12), (17) is a cubic cone with vertex at the point O:

$$2b_1b_2x^2z + (2b_1y - c_2z)(y - z/n)^2 = 0.$$

This cubic cone is intersected by the plane $x = 0$ in the arbitrarily chosen line (18), counted twice, and another line whose equations are

$$(19)\qquad x = 2b_1y - c_2z = 0.$$

This latter line is independent of the line (18). If the vertex $(0, 0, 0, 1)$ is on the line (19), then $c_2 = 0$. The system of reference is now determined except for the actual position of the unit point on the cone (7) of Moutard.

From the form of the coordinates of the vertex of the cone (15) of Moutard,

$$(20)\qquad (3b_1 + 2b_2\lambda, \lambda^2, 0, 0),$$

it follows that for any point, distinct from the point O and on the asymptotic tangent $y = z = 0$, there are two tangents of the surface S at O, associated with which the two cones of Moutard have a common vertex at the point. For convenience, we shall call the tangents *the associated tangents* of the point, and the point *the associated vertex* of the tangents. It should be

noted that one of the associated tangents of the vertex $(0, 0, 0, 1)$ has been chosen to be the edge $x = z = 0$.

The cone (7) intersects the surface S in a curve with a triple point at O. The residual triple-point tangent, other than the y-axis, is

$$z = b_1 x + b_2 y = 0. \tag{21}$$

If the unit point $(1, 1, 0)$ in the tangent plane $z = 0$ is taken on the harmonic conjugate of the tangent (21) with respect to the x- and y-axes, then $b_1 = b_2$. Moreover, with reference to the expression (20), we know that if the unit point $(1, 0, 0)$ on the asymptotic tangent $y = z = 0$ is the associated vertex of the tangent (21), then $b_2{}^2 = b_1$. Since $b_1 b_2 \neq 0$, we arrive at *the following canonical power series expansion of the surface S in the neighborhood of the parabolic point O*:

$$z = y^2 + x^3 + x^2 y + c_3 x^2 y^2 + \cdots.$$

Finally, we consider the equation of a general quadric having second order contact with the surface S at the point O, which may be obtained from the expansion (1):

$$z - a y^2 + (k_2 x + k_3 y + k_4 z) z = 0, \tag{22}$$

where k_2, k_3, k_4 are parameters. The quadric (22) intersects the surface S in a curve with a triple point at O, whose tangents have the equations

$$z = b_1 x^3 + b_2 x^2 y + (b_3 + a k_2) x y^2 + (b_4 + a k_3) y^3 = 0. \tag{23}$$

It is easy to show that there is a unique line,

$$z = 3 b_1 x + b_2 y = 0, \tag{24}$$

in which the three triple-point tangents (23) may coincide. Since the tangent at the point O to the parabolic curve or the locus of the parabolic points of the surface S has [10] the same equations as (24), we obtain the following conclusion:

At a parabolic point of a surface there is a unique tangent of Darboux, which coincides with the tangent to the parabolic curve on the surface.

3. Tangential section with a tacnode. This section is devoted to the case in which the tangential section of the surface S at the parabolic point O has an ordinary tacnode at O. The expansion of the surface S in the neighborhood of such a point O may be written in the form (1) with the

condition $b_1 = 0$. From equation (24) it follows immediately that the tangent at O to the parabolic curve of the surface S coincides with the coincident asymptotic tangents [11, 12].

For later use we find, from the expansion (1) and the condition $b_1 = 0$, the expansions of the two branches of the tangential section of the surface S at the point O, namely,

$$z = 0, \qquad y = Ax^2 + Bx^3 + Cx^4 + \cdots, \tag{25}$$

where the coefficients A, B, C satisfy the following equations

$$\begin{aligned} &aA^2 + b_2A + C_1 = 0, \\ &(2aA + b_2)B + b_3A^2 + c_2A + d_1 = 0, \\ &(2aA + b_2)C + aB^2 + 2b_3AB + C_2B + b_4A^3 + C_3A^2 + d_2A + e_1 = 0. \end{aligned} \tag{26}$$

By means of equations (3), (4), (5), (9), we may choose the system of reference so as to reduce the expansion (1) of the surface S at the point O to the following form:

$$\begin{aligned} z = y^2 + b_2x^2y + c_1x^4 + c_2x^3y + c_3x^2y^2 + d_1x^5 + d_2x^4y \\ + \cdots + d_6y^5 + \cdots. \end{aligned} \tag{27}$$

For the purpose of completing the determination of the system, let us recall an osculant of Su [13] associated with a singular point of a plane curve.

Let O be a singular point of order $m - 1$ ($\geqq 2$) of a plane curve C such that the tangent t_0 of the curve C at the point O has contact of order $m - 1$. Consider all the algebraic curves of order m in the plane of C every one of which has a multiple point of order $m - 1$ at a point M, not on t_0. Denote one of these curves by C_m and suppose that all the branches of C_m at M are tangent to the same line t at M, and further suppose that C_m has contact of order $m + 1$ with C at O. Then the tangent t intersects the tangent t_0 in a point O_{m+1} which is independent of the particular curve C_m being considered. If the expansion of the curve C in the neighborhood of the point O takes the form

$$\eta = a_0\xi^m + a_1\xi^{m+1} + \cdots \qquad (a_0 \neq 0), \tag{28}$$

the coordinates of the point O_{m+1} are

$$((m - 1)a_0/a_1, \quad 0). \tag{29}$$

From the series (27), (28) and the expression (29), it is easily seen

that a general plane through the asymptotic tangent cuts the surface S in a curve having O as a singular point of the third order, and that associated with any such plane section there is a fixed covariant point O_5 with the coordinates

$$(3c_1/d_1, \quad 0, \quad 0). \tag{30}$$

If the vertex $(0,1,0,0)$ is taken at the point (30), then $d_1 = 0$.

There is a nine-parameter family of cubic surfaces having third order contact with the surface S at the point O. The equation of a general one of these cubic surfaces is found, from the expansion (27), to be

$$\begin{aligned} z = y^2 + k_1xz + k_2yz + b_2x^2y - k_1xy^2 - k_2y^3 + k_3z^2 \\ + k_4xyz + k_5x^2z + k_6y^2z + k_7xz^2 + k_8yz^2 + k_9z^3, \end{aligned} \tag{31}$$

where the k's are parameters. The cubic surface (31) cuts the surface S in a curve with a quadruple point at O. The four quadruple-point tangents are

$$\begin{aligned} z = c_1x^4 + (c_2 - b_2k_1)x^3y + (c_3 - b_2k_2 - k_5)x^2y^2 \\ - k_4xy^3 - (k_3 + k_6)y^4 = 0. \end{aligned} \tag{32}$$

Furthermore, the intersection of the tangent plane $z = 0$ and the cubic surface (31) is a cubic curve which degenerates into the asymptotic tangent $y = z = 0$ and the conic

$$z = y + b_2x^2 - k_1xy - k_2y^2 = 0. \tag{33}$$

Among the cubic surfaces (31), we may therefore determine a five-parameter family satisfying the following two conditions:

(i) The y-axis is the polar line of the vertex $(0,1,0,0)$ with respect to the conic (33).

(ii) Three of the tangents (32) coincide with the y-axis.

For this five-parameter family, we have a common residual quadruple-point tangent of (32) with the equations

$$z = c_1x + c_2y = 0. \tag{34}$$

If we choose the vertex $(0,0,0,1)$ on the cone (15) of Moutard for the tangent (34), then $c_3 = (b_2c_2/c_1)^2$. Moreover, if the unit point $(1,1,0)$ is taken on the harmonic conjugate of the tangent (34) with respect to the x- and y-axes, then $c_2 = c_1$. Finally, from equations (11), (20) we observe

that for each non-asymptotic tangent in the tangent plane $z=0$ at the point O there is only one associated vertex on the asymptotic tangent. If we choose the unit point $(1,0,0)$ at the harmonic conjugate of the point O with respect to the vertex $(0,1,0,0)$ and the associated vertex of the unit tangent $z=y-x=0$, then $b_2=1$. Thus we are led to *the following canonical power series expansion of the surface S in the neighborhood of the point O*:

$$z=y^2+x^2y+c_1(x^4+x^3y)+x^2y^2+d_2x^4y+d_3x^3y^2+\cdots+d_6y^5+\cdots.$$

4. Tangential section with a simple inflexional tacnode. According as one or both branches of the tangential section T with a tacnode at a parabolic point O of a surface S have inflexions at O, we call the point O *a simple* or *double inflexional tacnode* of T. We shall discuss the first case in this section, and leave the other to the last section.

From the expansion (1), with the condition $b_1=0$, and the first of equations (26), it follows that the condition for the point O of T to be a simple inflexional tacnode is $c_1=0$. As in **3**, the expansion of the surface S at such a point O may also be reduced to the form (27) in which $c_1=0$. Any four-point conic of the branch of T, having O as an ordinary point, has the equations

$$z=y+b_2x^2+(1/b_2{}^2)(d_1-b_2c_2)xy+ky^2=0, \tag{35}$$

where k is a parameter. There is only one point on the asymptotic tangent $y=z=0$ such that its associated tangent coincides with its polar line with respect to the conic (35), namely, the point

$$(b_2c_2-d_1, 2b_2{}^2, 0, 0).$$

If we choose this point as the vertex $(0,1,0,0)$, then $d_1=b_2c_2$.

Among the cubic surfaces (31) we may determine an eight-parameter family such that two of the four quadruple-point tangents (32) coincide with the asymptotic tangent. For this family, $k_1=c_2/b_2$ and the polar line of the vertex $(0,1,0,0)$ with respect to the conic (33) is

$$z=c_2y-2b_2{}^2x=0. \tag{36}$$

The harmonic conjugate of the y-axis with respect to the x-axis and the tangent (36) has the equations

$$z=c_2y-b_2{}^2x=0. \tag{37}$$

If the vertex $(0,0,0,1)$ is taken on the cone (15) of Moutard for the tangent (37), then $c_3=0$.

The quartic surface, whose equation is

$$z = y^2 + b_2 x^2 y + c_2 x^3 y, \tag{38}$$

is completely characterized by the following properties: it has fourth order contact with the surface S at the point O; it has a unode of the third order at the point $(0, 0, 0, 1)$; and its uniplane is the plane $x_1 = 0$. The tangent plane $z = 0$ intersects the quartic surface (38) in a curve degenerating into the x-axis and the cubic curve

$$z = y + b_2 x^2 + c_2 x^3 = 0. \tag{39}$$

If the unit point $(1, 0, 0)$ is taken at the harmonic conjugate of the residual point of intersection, other than O, of the asymptotic tangent and the cubic curve (39) with respect to the point O and the vertex $(0, 1, 0, 0)$, then $c_2 = b_2$. Finally, by choosing the unit point $(1, 1, 0)$ on the tangent (37), we obtain *the following canonical power series expansion of the surface S in the neighborhood of the point O*:

$$z = y^2 + x^2 y + x^3 y + x^5 + d_2 x^4 y + d_3 x^3 y^2 + \cdots + d_6 y^5 + \cdots.$$

5. Tangential section with a symmetric tacnode. For an ordinary tacnode P of a plane curve C with tacnodal tangent t, the first polar curve Γ of a general point on t with respect to C has a double point at P with t as a tangent. If the curve Γ has an inflexion at the point P with t as the inflexional tangent, then the point P was called by Wölfing [14] and Segre [11] *a symmetric* or *harmonic tacnode* of the curve C. Segre [11] also showed that a necessary and sufficient condition for the parabolic curve of a surface S to have a double point at a parabolic point O is that the tangential section of the surface S at the point O have O as a symmetric tacnode.

The parabolic point O of the surface S with the general expansion (1) is a symmetric tacnode of the tangential section of S at O if and only if $b_1 = b_2 = 0$. Thus from (6), the expansion of the surface S at the point O for this case can be reduced to the following form

$$\begin{aligned} z = y^2 &+ c_1 x^4 + c_2 x^3 y + c_3 x^2 y^2 + c_4 x y^3 \\ &+ d_1 x^5 + d_2 x^4 y + \cdots + d_6 y^5 + \cdots. \end{aligned}$$

By means of the expression (20), it is obvious that all the tangents at the point O to the surface S have the vertex $(0, 1, 0, 0)$ as a common associated vertex. Making use of equation (25) and the first two of equations (26),

we may easily obtain the equations of the two polar lines of the vertex $(0,1,0,0)$ with respect to any four-point conics of the two branches of the tangential section T at the point O of the surface S, namely,

$$z = 4c_1\sqrt{-c_1}x + (c_2\sqrt{-c_1} \pm d_1)y = 0.$$

If the harmonic conjugate of the x-axis with respect to these two polar lines is the y-axis, then $c_2 = 0$ since c_1 does not vanish.

From equation (8), it follows that through the y-axis there is only one plane section of the surface S which has a sextactic point at O. If this plane is taken as the coordinate plane $x = 0$, then $d_6 = 0$.

We now consider the plane (12) passing through the general lines (18) in the plane $x = 0$. Observing the conditions $b_1 = b_2 = d_6 = 0$, calculating the series (13) to the fifth degree and then substituting the result for y in the left member of equation (14), we obtain

$$Q = (U/n)x^5 + \cdots, \tag{40}$$

the coefficient U being defined by the formula

$$\begin{aligned} U = d_1 + (d_2 - 4c_1/n)\lambda + (d_3 - 3c_2/n)\lambda^2 \\ + (d_4 - 2c_3/n)\lambda^3 + (d_5 - c_4/n)\lambda^4. \end{aligned} \tag{41}$$

Thus through a general line (18) in the plane $x = 0$ there are four plane sections of the surface S, each of which has a sextactic point at the point O. In particular we can determine a unique line through which only three such plane sections may be drawn. From the expression (41), we easily find the equation of this line to be

$$x = d_5z - c_4y = 0. \tag{42}$$

If the vertex $(0,0,0,1)$ is on this line, then $d_5 = 0$.

If we take the harmonic conjugate of the point O with respect to the vertex $(0,1,0,0)$ and the covariant point (30) as the unit point $(1,0,0)$, then $d_1 = 6c_1$. Finally, if the unit point $(1,1,0)$ is chosen on the polar line of the unit point $(1,0,0)$ with respect to any four-point conics of any one branch of the tangential section T at the point O of the surface S, then $c_1 = -1$.

Hence we obtain *the following canonical power series expansion of the surface S in the neighborhood of the point O*:

$$z = y^2 - x^4 + c_3x^2y^2 + c_4xy^3 - 6x^5 + d_2x^4y + d_3x^3y^2 + d_4x^2y^3 + \cdots.$$

6. Tangential section with a double inflexional tacnode. Finally, we shall consider the case in which the tangential section T of the surface S at a parabolic point O has a double inflexional tacnode at O. From the general expansion (1) and equations (25), (26) it is easily seen that the conditions for this case are $b_1 = b_2 = c_1 = d_1 = 0$. The expansion (6) is then reduced to the form

$$(43) \qquad z = y^2 + c_2x^3y + c_3x^2y^2 + c_4xy^3 + d_2x^4y + d_3x^3y^2 + \cdots + d_6y^5 + \cdots.$$

As in the preceding section, all the tangents at the point O of the surface S have the vertex $(0, 1, 0, 0)$ as a common associated vertex.

By means of equation (15) and the expansion (43), we see that the cone of Moutard for a general tangent t_λ (11) intersects the surface S in a curve with a quadruple point at O, and that two of the quadruple-point tangents are the asymptotic tangent and the tangent t_λ. In particular, we can determine two tangents t_λ such that for each of them one of the two residual quadruple-point tangents of the intersection further coincides with it. If we choose the harmonic conjugate of the asymptotic tangent with respect to these two tangents to be the y-axis, then $c_3 = 0$.

From equation (8), it follows that through the y-axis there is only one plane section of the surface S which has a sextactic point at O. If this plane is taken as the coordinate plane $x = 0$, then $d_6 = 0$.

By using equations (14), (40), (41) we can easily show that through a general line (18) in the plane $x = 0$ there are three plane sections of the surface S each of which has a sextactic point at O. In particular, there is a unique line (42) through which only two such plane sections can be drawn. If the vertex $(0, 0, 0, 1)$ is on this line (42), then $d_5 = 0$.

The cone (7) cuts the surface S in a curve with a quadruple point at O. The residual quadruple-point tangents, other than the x- and y-axes, are

$$(44) \qquad z = c_2x^2 + c_4y^2 = 0.$$

If the unit point $(1, 1, 0)$ in the tangent plane $z = 0$ is on one of the two tangents which are harmonic conjugate with respect to the pair of the tangents (44) and the pair of the x-, y-axes, then $c_4 = c_2$.

Finally, the z-axis intersects the cone (15) of Moutard for the unit tangent $z = y - x = 0$ in two points P_1 and P_2. If the unit point $(0, 0, 1)$ on the z-axis is chosen at the harmonic conjugate of the point O with respect to the vertex $(0, 0, 0, 1)$ and one of the two points P_1 and P_2, then $c_2 = 1$.

Thus we reach *the following canonical power series expansion of the surface S in the neighborhood of the point O*:

$$z = y^2 + x^3y + xy^3 + d_2x^4y + d_3x^3y^2 + d_4x^2y^3 + \cdots.$$

MICHIGAN STATE COLLEGE.

BIBLIOGRAPHY

1. E. Bompiani, "Per lo studio proiettive-differenziale delle singolarità," *Bollettino della Unione di Matematica Italiana*, vol. 5 (1926), pp. 118-120.

2. ———, "Geometria proiettiva di elementi differenziali," *Annali di Matematica*, (4), vol. 22 (1942), pp. 1-32.

3. ———, "Analisi dei flessi di specie superiore delle curve piane," *Bollettino della Unione di Matematica Italiana*, (2), vol. 5 (1943), pp. 156-168.

4. T. L. Downs, "Asymptotic and principal directions at a planar point of a surface," *Duke Mathematical Journal*, vol. 1 (1935), pp. 316-327.

5. V. G. Grove, "Differential geometry of a surface at a planar point," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 155-166.

6. ———, "Geometry of a surface in the neighborhood of a spine," *Transactions of the American Mathematical Society*, vol. 45 (1939), pp. 256-264.

7. E. P. Lane, "The neighborhood of a sextactic point on a plane curve," *Duke Mathematical Journal*, vol. 1 (1935), pp. 287-292.

8. I. Popa, "Geometria proiettivo-differenziale delle singolarità delle curve piane," *Rendiconti dei Lincei*, (6), vol. 25 (1937), pp. 220-222.

9. ———, "Geometria proiettivo-differenziale delle singolarità: Il punto di flesso di una superficie qualunque," *Rendiconti dei Lincei*, (6), vol. 26 (1937), pp. 147-150.

10. ———, "Contributo allo studio proiettivo-differenziale delle singolarità. Punto di flesso e punto parabolico delle superficie," *Rendiconti del Seminario Matematico della R. Università di Roma*, (4), 2 (1938), pp. 136-155.

11. C. Segre, "Su alcuni punti singolari delle curve algebriche, e sulla linea parabolica di una superficie," *Rendiconti dei Lincei*, (5), vol. 6 (1897), pp. 168-175.

12. V. Strazzeri, "Sulla linea dei punti parabolici di una superficie," *Rendiconti del Circolo Matematico di Palermo*, vol. 41 (1916), pp. 54-58.

13. B. Su, "An extension of Bompiani's osculants for a plane curve with a singular point," *Tôhoku Mathematical Journal*, vol. 45 (1938), pp. 239-244.

14. E. Wölffing, "Ueber die Hesse'sche Covariante einer ganzen rationalen Function von ternären Formen," *Mathematische Annalen*, vol. 36 (1889-90), pp. 97-120.

THE DIFFERENTIAL EQUATION $y' = f(y)$.*

By SYLVAN WALLACH.

1. Let $f(y)$ be a continuous, real-valued function on the interval $c \leqq y \leqq d$. It is known [1] that if

$$(1) \qquad f(y) > 0 \text{ for } c < y < d,$$

then the differential equation and the initial condition

$$(2) \qquad y' = f(y), \qquad y(0) = c$$

have a solution $y = y(x) \not\equiv c$ on some interval $0 \leqq x \leqq b$ if and only if

$$(3) \qquad \int_{c+0} dy/f(y) < \infty.$$

It is the object of this note to describe the general situation, in which condition (1) is not *assumed.* The solution of the general problem depends on tools less elementary than one might suspect. It is true that all solutions prove to be monotone, and are, of course, absolutely continuous. The problem depends, however, on the inversion of a monotone function, and the inversion leads, in general, to monotone functions which are not absolutely continuous. For this reason the description of the general situation will have to involve Lebesgue's decomposition of a monotone function into absolutely continuous, continuous but singular, and purely discontinuous components.

As to the mere existence of solutions, the situation is as follows:

(I) *Let $f(y)$ be a real-valued, continuous function on the interval $c \leqq y \leqq d$. In order that (2) possess a solution $y = y(x) \not\equiv c$ on some interval $0 \leqq x \leqq b$, it is necessary and sufficient that there exists a number d^* satisfying*

$$(4) \qquad f(y) \geqq 0 \textit{ for } c \leqq y \leqq d^* \qquad (c < d^* \leqq d)$$

and

$$(5) \qquad \int_c^{d^*} ds/f(s) < \infty.$$

* Received June 7, 1947.

If (4) *and* (5) *are satisfied, a strictly monotone solution* $y = y(x)$ *results by inversion of the quadrature*

$$x(y) = \int_c^y ds/f(s)\,; \tag{5 bis}$$

and this solution is an extreme solution.

If $f(c) \neq 0$, the sufficiency of the conditions is trivial and the solution of (2) is unique. If the conditions are satisfied and $f(c) = 0$, any solution $y = y_1(x)$ of (2) satisfies

$$c \leqq y_1(x) \leqq y(x), \qquad (0 \leqq x \leqq b),$$

where $y = y(x)$ is that solution of (2) obtained by the inversion of the quadrature (5 bis). It is in this sense in which the solution supplied by (5 bis) is called "extreme."

The content of Theorem (I), when (1) is not satisfied, will be illustrated by an example to the following effect:

(II) *There exist real-valued, continuous functions* $f(y)$, *where* $c \leqq y \leqq d$, *with the properties that* (4) *and* (5) *are satisfied* (*with* $d = d^*$) *and that* (2) *possesses strictly monotone solutions distinct from the solution obtained by the inversion of the quadrature* (5 bis).

It may be remarked that (II) is equivalent to the following fact concerning monotone functions:

(III) *There exist two monotone increasing functions* $y_1(x)$, $y_2(x)$, *where* $0 \leqq x \leqq 1$, *with the properties that* $y_1(0) = y_2(0) = 0$, *the respective inverse, functions* $x_1(y)$, $x_2(y)$ *possess continuous derivatives, and* $x'_1(y) \equiv x'_2(y)$, *where* $0 \leqq y \leqq \min(y_1(1), y_2(1))$ *but* $y_1(x) \not\equiv y_2(x)$.

It will follow that, if conditions (4) and (5) of (I) are satisfied, the non-constant solutions of (2) can be described as follows:

(IV) *Let* $f(y)$ *be a real-valued, continuous function on the interval* $c \leqq y \leqq d$, *and let* $f(y)$ *satisfy conditions* (4) *and* (5). *Define*

$$x(y) = x_1(y) + x_2(y) + x_3(y), \text{ for } c \leqq y \leqq d^*, \qquad (c < d^* \leqq d), \tag{6}$$

where $x_1(y)$ *is the quadrature* (5); $x_2(y)$ *is non-decreasing, continuous, singular, and constant in each interval where* $f(y) > 0$; *and* $x_3(y)$ *is a*

purely discontinuous, non-decreasing function whose jumps occur only where $f(y)$ *vanishes.*

The non-constant solutions of (2) *are provided by that inverse of* (6) *defined as follows:*

$$y(x) = \text{l. u. b. } y \textit{ for } x(y) \leqq x, \tag{7}$$

where $0 \leqq x \leqq x(d^*)$.

2. In order to prove (I), some elementary facts will first be proved.

(i) *If* $f(y)$ *is a continuous function on* $c \leqq y \leqq d$ *and* $y = y(x)$ *is a solution of* (2) *for* $0 \leqq x \leqq b$, *then* $y(x)$ *is monotone for* $0 \leqq x \leqq b$.

For, if the solution $y(x)$ were not monotone, there would exist two numbers $x_1 \neq x_2$ on the interval $(0, b)$ such that $y(x_1) = y(x_2)$ and $y'(x_1) > 0$, $y'(x_2) < 0$. This contradicts (2), since (2) implies that y' is a single-valued function of y.

(ii) *If* $f(y)$ *is a continuous function on* $c \leqq y \leqq d$ *and if* $f(y)$ *changes sign on every interval* $(c, c + \epsilon)$, *then* $y = y(x) \equiv c$ *is the only solution of* (2).

The continuity of $f(y)$ implies that $f(c) = 0$ and, therefore, that $y(x) \equiv c$ is a solution of (2). Suppose $y = y(x)$ were a solution of (2) for $0 \leqq x \leqq b$ and that $y(x) \not\equiv c$. Then for some $x = x_1$, where $0 < x_1 \leqq b$, the function value $y(x_1) > c$. But $f(y)$ changes sign in the interval $(c, y(x_1))$, which means, by (2), that $y = y(x)$ is not monotone on $0 \leqq x \leqq x_1$. This contradicts (i) and so (ii) follows.

3. The proof of the necessity of the conditions (4) and (5) in (I) will now be given. It will, therefore, be supposed that $y = y(x) \not\equiv c$, where $0 \leqq x \leqq b$, is a solution of (2). The statement (ii) implies that there exists a number d^*, where $c < d^* \leqq d$, such that $f(y)$ does not change sign on (c, d^*). Consequently, (4) holds. For otherwise

$$f(y) \leqq 0 \text{ for } c \leqq y \leqq d^*,$$

which, by (2) and (i), means that $y(x)$ is non-increasing on $0 \leqq x \leqq b$. Hence, $y(x) \not\equiv c$ implies $y(x) < c$ for some value of x. But this contradicts the assumption that $f(y)$ is defined only for $c \leqq y \leqq d$.

It will now be shown that (5) holds if $d^* = y(b)$. Since $y(x) \not\equiv c$, this d^* satisfies $c < d^* \leqq d$. It is known that a monotone function $y = y(x)$ maps the set of x-values on which the derivative $y'(x)$ vanishes on a y-set of

measure zero. Hence, by (2), the zeros of $f(y)$, where $0 \leqq y \leqq d^*$, form a y-set, F, of measure zero. The continuity of $f(y)$ implies that F is a closed set. Let E denote the open set which is the complement of F in (c, d^*). Let $E = \Sigma(c_n, d_n)$ be the decomposition of E into disjoint open intervals and let the numbers a_n, b_n be defined by $y(a_n) = c_n$, $y(b_n) = d_n$. It follows, from the definition of the open set E, that $f(y) > 0$ for $c_n < y < d_n$, and from the theorem quoted in the first paragraph of **1**, that the integral

$$\int_{c_n}^{d_n} ds/f(s)$$

converges and has the value $b_n - a_n$. Hence,

$$\int_E ds/f(s) = \Sigma(b_n - a_n).$$

Since the disjoint open intervals (a_n, b_n) are contained in $(0, b)$ and since the complement F of E in (c, d^*) is a zero set, it follows that

(8) $$\int_c^{d^*} ds/f(s) \leqq b < \infty.$$

This completes the proof of the necessity of (5).

4. The proof of the sufficiency of the conditions (4) and (5) in (I) will now be given. These conditions imply that the y-set on which $f(y)$ vanishes is a zero set on (c, d^*). Hence,

(9) $$x(y) = \int_c^y ds/f(s)$$

is a strictly increasing function of y and so possesses an inverse function $y = y(x)$, where $0 \leqq x \leqq b$ and $b = x(d^*)$. Then (9), (4) and the continuity of $f(y)$ imply that

$$y'(x) = 1/x'(y) = f(y(x)), \qquad (0 \leqq x \leqq b),$$

where it is understood that $1/\infty = 0$.

5. In order to complete the proof of (I), it remains to be shown that, if (4) and (5) are satisfied, the function $y = y(x)$ obtained by the inversion of (9) is an extreme solution. To this end, let $y = y_1(x)$ be any solution of (2) on some interval $0 \leqq x \leqq b_1$. It will be proved that

(10) $$c \leqq y_1(x) \leqq y(x) \text{ for } 0 \leqq x \leqq b_2 = \min(b, b_1).$$

The first inequality in (10) is clear. It follows from (i) that $y_1(r)$ is monotone. If $y_1(x) \equiv c$, then (10) is obvious. Hence, suppose $y_1(b_2) = d_2 > c$. Let $x_1(y)$, where $c \leqq y \leqq d_2$, be an inverse (possibly discontinuous) of $y_1(x)$. It is clear from the arguments employed in **3** that

$$x_1(y) \geqq \int_c^y ds/f(s) = x(y);$$

cf. (8). This relation between the inverse functions $x_1(y)$ and $x(y)$ shows that the second inequality in (10) holds. This completes the proof of (I).

6. In order to prove (II), let T be a nowhere dense, closed set on $0 \leqq x \leqq 1$ such that the common part of T with every interval $(0, x)$ contained in $(0, 1)$ has positive measure. Let S be the complement of T in $0 \leqq x \leqq 1$. Let $g(x)$ be a continuous function on $0 \leqq x \leqq 1$ such that $g(x) = 0$ if x is in T and $g(x) > 0$ if x is in S (that such a $g(x)$ exists, follows by taking the absolute value of Volterra's well-known example). Define $y = y_1(x)$ by

$$y_1(x) = \int_0^x g(t)\,dt, \qquad (0 \leqq x \leqq 1). \tag{11}$$

Since the set S is dense in $(0, 1)$ and since $g(x) > 0$ on S, it follows that $y_1(x)$ is monotone increasing for $0 \leqq x \leqq 1$. Let $x_1(y)$ denote the inverse function of $y_1(x)$ for $0 \leqq y \leqq d$, where $d = y_1(1)$. Let $f(y)$ be defined by

$$f(y) = y'(x_1(y)) \text{ for } 0 \leqq y \leqq d. \tag{12}$$

Then (11) is a monotone increasing solution of (2), where $c = 0$. Hence, by (I), the function (12) satisfies conditions (4) and (5). On the other hand, the image of the set T under the mapping $y = y_1(x)$ is a zero set, since $y'_1(x) = g(x) = 0$ on T. Hence, the inverse function $x = x_1(y)$ maps a zero set onto the x-set T of positive measure and, therefore, is not absolutely continuous. Consequently the function (11) is not the inverse of the quadrature (9). This completes the proof of (II).

7. In order to prove (IV), it will be shown, first, that the function $y(x)$, defined in (7) is a solution of (2). But this is obvious from the argument in **4**, since $x_1(y)$ can be identified with (9), and the derivates of $x_2(y)$ and $x_3(y)$ are bounded from below.

Conversely, let

$$y^*(x), \text{ for } 0 \leqq x \leqq b_1, \tag{13}$$

be any non-constant solution of (2), where $f(y)$ is defined in (IV). In view of (i) and (10), $y^*(x)$ is non-decreasing. Let $x^*(y)$ be the inverse of $y^*(x)$ defined as follows:

$$x^*(y) = \text{l.u.b.}\, x \text{ for } y^*(x) \leqq y, \tag{14}$$

where $c \leqq y \leqq y(b_1)$. The function $x^*(y)$ is monotone increasing. According to the decomposition theorem of Lebesgue

$$x^*(y) = x^*_1(y) + x^*_2(y) + x^*_3(y), \tag{15}$$

where the three functions on the right in (15) are non-decreasing; $x^*_1(y)$ is absolutely continuous; $x^*_2(y)$ is continuous and purely singular; and $x^*_3(y)$ is purely discontinuous. The total variation of $x^*_2(y) + x^*_3(y)$ clearly occurs on the y-set where $y'(x) = 0$, hence where $f(y) = 0$. Therefore, $x^*(y)$ satisfies the conditions imposed on $x(y)$ in (IV). Since (13) is obtained from (15) in the same way that (7) is obtained from (6), this completes the proof of (IV).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

[1] E. Kamke, *Differentialgleichungen reeller Funktionen,* Leipzig (1930); cf. pp. 7-21.

THE NEIGHBORHOOD OF AN UNDULATION POINT ON A SPACE CURVE.*

By Alice T. Schafer.

In projective differential geometry the neighborhood of an ordinary point on a curve in three-space has already been studied [3; 61-62].[1] In a previous paper by the author a study was made of the neighborhood of two singular points on a space curve: the inflexion point and the planar point [4]. The next singular point which naturally presents itself for study is the undulation point, which is defined as a point at which the tangent to the curve has precisely four-point contact with the curve instead of the usual two-point contact. For this reason the point is classed as a singular point although the curve may be represented in the neighborhood of the point by power series as is the case in the neighborhood of an ordinary point.

In **1**, the projective coordinate system, consisting of a tetrahedron of reference and a unit point, is chosen so as to give canonical power-series expansions for the curve in the neighborhood of an undulation point. These series are then used to study properties of the curve in the neighborhood of the singular point; **2** is devoted to a study of surfaces osculating the curve at the undulation point; in **3** plane sections of the tangent developable in the neighborhood of the singularity are investigated and in **4** a similar study is made of the projections of the curve from the faces of the tetrahedron of reference.

Throughout the paper homogeneous and nonhomogeneous coordinates are used interchangeably. A point is represented in nonhomogeneous coordinates by x, y, z and in homogeneous coordinates by x_1, x_2, x_3, x_4. The two systems of coordinates are connected by the relations $x = x_2/x_1$, $y = x_3/x_1$, $z = x_4/x_1$.

1. Canonical power-series expansions. In the neighborhood of any point on an analytic curve in three-space two of the coordinates may be expressed as power series in terms of the third. Let the coordinates of a variable point on the curve be represented by (x, y, z) and the point O around which the expansions are to be taken by (c_0, a_0, b_0); then the expansions will be

* Received April 15, 1947.

[1] Numbers in brackets refer to the works listed in the bibliography at the end of the paper.

$$
(1)\qquad \begin{aligned} y &= a_0 + a_1(x - c_0) + a_2(x - c_0)^2 + \cdots, \\ z &= b_0 + b_1(x - c_0) + b_2(x - c_0)^2 + \cdots, \end{aligned}
$$

where a_i and b_i $(i = 0, 1, 2, \cdots)$ are constants. We shall choose the point O as the undulation point we are to study and then prove the following:

THEOREM 1.1. *By a suitable choice of the projective coordinate system in the neighborhood of an undulation point on a space curve, the power-series expansions representing the curve in the neighborhood of the undulation point can be reduced to the following form*

$$
(2)\qquad \begin{aligned} y &= ax^4 - (3/2)ax^5 + a_6x^6 + a_7x^7 + a_8x^8 + a_9x^9 + a_{10}x^{10} \\ &\quad + a_{11}x^{11} + a_{12}x^{12} + a_{13}x^{13} + \cdots, \\ z &= ax^5 - 2ax^6 + (a_6 + 3a/4)x^7 + (a_7 - a_6/2 + a/8)x^8 \\ &\quad + (a_8 - a_7/2 - 3a/16)x^9 + (a_9 - a_8/2 + a_6/8 - a/16)x^{10} \\ &\quad + (a_{10} - a_9/2 + a_7/8 + 3a/32)x^{11} + (a_{11} - a_{10}/2 + a_8/8 \\ &\quad - a_6/16 + 5a/128)x^{12} + b_{13}x^{13} + \cdots \qquad (a \neq 0). \end{aligned}
$$

There are fifteen arbitrary parameters involved in the determination of the projective coordinate system and therefore fifteen conditions which may be imposed on the constants in the equations (1) in deducing canonical power-series expansions for the space curve C in the neighborhood of the undulation point O. In order to derive equations (2) from equations (1), we first take the vertex $(1, 0, 0, 0)$ of the projective tetrahedron of reference to be at the point $O(c_0, a_0, b_0)$, thus making $a_0 = b_0 = c_0 = 0$. The tangent to C at O now has the equations $y = a_1x$, $z = b_1x$; we shall place the edge $y = z = 0$ of the tetrahedron along this tangent line, thus making $a_1 = b_1 = 0$. Since O is an undulation point, $a_2 = a_3 = b_2 = b_3 = 0$ and not both a_4 and b_4 equal zero if O is not to be a singularity of higher order. We shall assume this to be the case.

Since O is an undulation point any plane through the tangent to C at O intersects C in four consecutive points at O and so the osculating plane of C at O is indeterminate. However, the plane $a_4z - b_4y = 0$ has five-point contact with C at O and we shall call this plane the hyperosculating plane of C at O. We choose the face $z = 0$ of the tetrahedron to coincide with this plane, making $b_4 = 0$ and $a_4 \neq 0$. For convenience we replace a_4 by a, b_5 by b. Of the fifteen conditions which we are at liberty to impose on the coordinate system, six have already been determined. We now proceed to impose the remaining nine conditions.

In order to do this we consider cubic surfaces having contact of various orders with the curve C at the point O. To find the equations of such

surfaces, we write the most general equation of the third degree in x, y, and z and demand that the expansions (1) for C satisfy this equation through terms in x to the power desired. First we demand that equations (1) satisfy the general cubic equation through terms in x^3, thus getting the equation of the 15-parameter family of cubics having 4-point contact with C at O. It can easily be shown that the tangent to C at O, $y=z=0$, lies on all these cubic surfaces. Among these surfaces there is a 9-parameter family, each of which has at least 8-point contact with C at O, on which the tangent to C at O is a double line, making these surfaces ruled surfaces. Let us consider the unique one of these ruled surfaces osculating the curve at O and having 17-point contact with C at O. Its equation in homogeneous coordinates is

$$
\begin{aligned}
(3)\qquad & Bx_3^3+Dx_4^3+Fx_2x_3^2+Gx_3^2x_4+Hx_2x_4^2+Ix_3x_4^2+Jx_2x_3x_4 \\
& \quad +Mx_1x_4^2+Rx_1x_3x_4=0,
\end{aligned}
$$

where the capital letters are homogeneous parameters representing functions of the coefficients of x in equations (1) and where not both M and R are zero for a non-composite osculating cubic surface. The coefficients in (3) are found by demanding that the series (1) for y and z satisfy equation (3) through terms in x^{16}. Since these coefficients are long, we shall omit writing them down until later.

Choose the vertices $(0,0,1,0)$ and $(0,0,0,1)$ at points on the surface represented by equation (3). This choice makes $B=D=0$. The equation of the tangent plane to (3) at $(0,0,1,0)$ is $Fx_2+Gx_4=0$. Choose the face $x_2=0$ along this plane, making $G=0$, $F\neq 0$. The equation of the tangent plane to the surface at $(0,0,0,1)$ is $Mx_1+Hx_2+Ix_3=0$. Take the face $x_1=0$ along this plane; then $H=I=0$ and $M\neq 0$.

The polar quadric of the ruled surface represented by equation (3) with respect to $(0,1,0,0)$ consists of two planes $x_3=0$ and $Fx_3+Jx_4=0$. Put the unit point $(1,1,1,1)$ on the second of these two planes making $F=-J$. The polar quadric of the cubic surface with respect to $(0,0,1,0)$ has the equation $Rx_1x_4+2Fx_2x_3+Jx_2x_4=0$; put the unit point on this quadric. Then $F=-R$; $R\neq 0$ since $F\neq 0$. It follows that $J=R$.

The polar quadric of $(0,0,0,1)$ with respect to the cubic surface (3) has the equation $Rx_1x_3+2Mx_1x_4+Rx_2x_3=0$ and the equation of the tangent plane to this quadric at the point $O(1,0,0,0)$ has the equation $Rx_3+2Mx_4=0$. If we put the unit point on this plane, then $M=-R/2$. The equation (3) now takes the form

$$
(4)\qquad 2x_2x_3^2-2x_2x_3x_4+x_1x_4^2-2x_1x_3x_4=0.
$$

As was stated previously, the coefficients in (3) are found by demanding that the series (1) satisfy (3) through terms in x^{16}. These coefficients are now given with the simplifications the above conditions will make. First $F = -bR/a$. Since we have $F = -R$, $R \neq 0$, $a = b$. Similarly $J = -M - R(b_6 - a_5)/a$; however, $J = R$, $M = -R/2$, and hence $b_6 = a_5 - a/2$. Next $H = (b_7 - a_6 + a_5/2)/a$, but $H = 0$, so that $b_7 = a_6 - a_5/2$. Similarly $B = (b_8 - a_7 + a_6/2 - a/8)/a^2 = 0$, so that $b_8 = a_7 - a_6/2 + a/8$. In the same way, $G = (b_9 - a_8 + a_7/2 - a_5/8)/a^2$, but $G = 0$, so that $b_9 = a_8 - a_7/2 + a_5/8$. Also $I = (b_{10} - a_9 + a_8/2 - a_6/8 + a/16)/a^2 = 0$ so that $b_{10} = a_9 - a_8/2 + a_6/8 - a/16$. In addition, $D = (b_{11} - a_{10} + a_9/2 - a_7/8 + a_5/16)/a^2 = 0$ so that $b_{11} = a_{10} - a_9/2 + a_7/8 - a_5/16$. From $2M + R = 0$, we get $b_{12} = a_{11} - a_{10}/2 + a_8/8 - a_6/16 + 5a/128$. Equations (1) now have the form

$$
\begin{aligned}
y &= ax^4 + a_5x^5 + a_6x^6 + a_7x^7 + a_8x^8 + a_9x^9 + a_{10}x^{10} \\
&\quad + a_{11}x^{11} + a_{12}x^{12} + a_{13}x^{13} + \cdots, \\
z &= ax^5 + b_6x^6 + b_7x^7 + b_8x^8 + b_9x^9 + b_{10}x^{10} + b_{11}x^{11} \\
&\quad + b_{12}x^{12} + b_{13}x^{13} + \cdots \qquad (a \neq 0)
\end{aligned}
\tag{5}
$$

where

$$
b_n = a_{n-1} - a_{n-2}/2 + a_{n-4}/8 - a_{n-6}/16 + 5a_{n-8}/128, \qquad (n = 6, \cdots, 12). \tag{6}
$$

Of the fifteen conditions we are at liberty to impose on the coordinate system, we have already imposed fourteen. We have yet to locate the face $x_3 = 0$ of the tetrahedron. In order to do this, we consider the quartic cone with vertex at O and osculating C at O. Its equation is

$$
y^4 - axz^3 + Ay^3z + By^2z^2 + Dyz^3 + Ez^4 = 0, \tag{7}
$$

where

$$
\begin{aligned}
A &= -\{a_5 + 3a/2\}/a, \qquad B = \{a_5^2 + aa_5 - aa_6\}/a^2, \\
D &= \{a^3/4 + (a^2a_5)/4 + (a^2a_6)/2 - a^2a_7 - (aa_5^2)/2 + 2aa_5a_6 - a_5^3\}/a^3, \\
E &= \{(3a^4)/16 + (a^3a_6)/4 - a^3a_8 - (a^2a_5^2)/4 + a^2a_6^2 + 2a^2a_5a_7 \\
&\quad - 3aa_5^2a_6 + a_5^4\}/a^4.
\end{aligned}
$$

The polar plane of the vertex $(0, 0, 1, 0)$ with respect to the cone (7) has the equation $4x_3 + Ax_4 = 0$. If the face $x_3 = 0$ is chosen to coincide with this plane, then $A = 0$ and $a_5 = -(3a)/2$. Now the coordinate system has been uniquely determined and with this condition on a_5 substituted in equations (5), Theorem 1.1 is proved.

2. Osculants of the curve. In this section equations (2) will be used to study properties of osculants of an analytic space curve C in the neighborhood of an undulation point O on the curve. It would be natural to consider first the quadric cone with vertex at O osculating C at O. To find the equation of this cone it is sufficient to write the most general homogeneous equation of the second degree in x, y, and z and to demand that the equation be satisfied in x by equations (2) up to and including terms in x^9. When this is done it can be seen that the equation of the cone is $z^2 = 0$. It would then be natural to consider the general quadric surface osculating C at O. By writing the most general equation of the second degree in x, y, and z and demanding that it be satisfied by equations (2) through terms in x^9, it can be seen that the equation of the osculating quadric is also $z^2 = 0$. These results give the following:

THEOREM 2.1. *The only* 10-*point cone with vertex at an undulation point O on a space curve C and osculating C at O is the hyperosculating plane of C at O counted twice. The* 10-*point osculating quadric of C at O is also the hyperosculating plane counted twice.*

We turn our attention, therefore, to cubic surfaces osculating C at O. First, by the method used above, we find the equation of the osculating cubic cone, with vertex at O, of C at. O to be $z^3 = 0$ which gives the following:

THEOREM 2.2. *The* 15-*point osculating cubic cone, with vertex at the undulation point O, of a space curve C at O is the hyperosculating plane of C at O counted three times.*

Let us then consider arbitrary cubic surfaces osculating C at O. If we write the most general equation of a cubic surface and demand that it intersect C at O in four consecutive points, the equation of such a surface will be of the form:

$$\begin{aligned}(8)\qquad & Ay^3 + Bz^3 + Dx^2y + Ex^2z + Fxy^2 + Gy^2z + Hxz^2 + Iyz^2 + Jxyz \\ & + Ly^2 + Mz^2 + Nxy + Qxz + Ryz + Sy + Tz = 0,\end{aligned}$$

where the capital letters denote homogeneous parameters. We note a property of this surface.

THEOREM 2.3. *The tangent line of a space curve C at an undulation point O lies on every one of the* 15-*parameter family of cubic surfaces having* 4-*point contact with C at O.*

Among this 15-parameter family of cubic surfaces is a 9-parameter family

of ruled surfaces on which the tangent of C at O is the double line. The equation of this family is given by equation (8) with $D=E=N=Q=S=T=0$. In this 9-parameter family of cubics is a surface having 17-point contact with C at O and osculating C at O. Its equation is (4). It was this surface which was used in developing the canonical expansions (2) for C in the neighborhood of O, and it is this surface which we discuss now.

As is known from the theory of projective differential geometry, there are only two projectively inequivalent cubic ruled surfaces [2; 145]. One of these has distinct directrices; a canonical form of its equation is $x_3x_1{}^2-x_4x_2{}^2=0$. The other, with directrices coinciding, is called Cayley's cubic scroll; a canonical form of its equation is $x_2{}^3+x_1(x_1x_3+x_2x_4)=0$. We state the following theorem showing that the surface (4) is projectively equivalent to the first of these two.

THEOREM 2.4. *The cubic ruled surface* (4) *is projectively equivalent to the cubic ruled surface whose canonical form is given by the equation* $x'_3x'_1{}^2-x'_4x'_2{}^2=0$. *The projective transformation taking* (4) *into this canonical form is*

$$x_1=-i(x'_3+x'_4)/2, \qquad x_2=(x'_3-x'_4)/2,$$
$$x_3=x'_1(1-i)/2+x'_2(1+i)/2, \qquad x_4=i(x'_2-x'_1), \qquad i=\sqrt{-1}.$$

The equations of the one-parameter family of rulings on the surface (4) are

$$kx_1(k+2h)+2hx_2(k+h)=0, \qquad kx_3+hx_4=0,$$

where h and k are homogeneous parameters. Among the special lines on this surface, other than the double line (the tangent $x_3=x_4=0$ of C at O), are the edges $x_2=x_4=0$ $(k=0)$, $x_1=x_3=0$ $(h=0)$, $x_1=x_2=0$ of the tetrahedron of reference, the line $x_4-2x_3=0$ $(2h=-k)$ in the face $x_2=0$ and the line $x_3-x_4=0$ $(h=-k)$ in the face $x_1=0$.

All four vertices of the tetrahedron of reference are on the cubic ruled surface (4), but the unit point $(1,1,1,1)$ is not. At the points $(1,0,0,0)$ and $(0,1,0,0)$ the tangent planes to the surface are indeterminate since the line $x_3=x_4=0$ joining them is a double line on the surface. At the vertex $(0,0,1,0)$ the tangent plane is the face $x_2=0$ of the tetrahedron; at the vertex $(0,0,0,1)$ the tangent plane is the face $x_1=0$. The equation of the polar plane of the unit point with respect to the surface (4) is $x_1+2x_4=0$.

The polar quadric of the vertex $(1,0,0,0)$ with respect to the cubic surface (4) is degenerate, being composed of the hyperosculating plane, $x_4=0$, of C at O and the plane $x_4-2x_3=0$. Similarly the polar quadric

of the vertex $(0,1,0,0)$ is composed of two planes: the face $x_3=0$ and the plane $x_3=x_4$ through the unit point. The equation of the polar quadric of the vertex $(0,0,1,0)$ is $x_4(x_1+x_2)-2x_2x_3=0$ and the equation of the polar quadric of $(0,0,0,1)$ is $x_3(x_1+x_2)-x_1x_4=0$. The polar quadric of the unit point has the equation $x_1x_3-x_2x_3+x_2x_4-x_3^2+2x_3x_4-x_4^2/2=0$.

As is known, the Hessian of a surface [1; 313] is the locus of all the points whose polar quadrics with respect to the surface are cones. In the case of the cubic ruled surface (4) the Hessian has the equation

$$\{2x_3-x_4(1+i)\}^2\{2x_3-x_4(1-i)\}^2=0.$$

We have the following:

THEOREM 2.5. *The Hessian of the cubic ruled surface represented by equation* (4) *consists of two planes through the tangent of C at O, each counted twice. The intersection of the cubic surface and the Hessian is the tangent of C at O counted eight times and the two generators of the cubic ruled surface*

$$2x_3-x_4(1+i)=0, \qquad ix_1+x_2=0,$$

and

$$2x_3-x_4(1-i)=0, \qquad ix_1-x_2=0,$$

each counted twice.

If equations (2) are substituted in equation (4), we have

$$2xy^2-2xyz+z^2-2yz=(-2ab_{13}+2aa_{12}-aa_{11}+aa_9/4 \\ -aa_7/8-15a^2/128)x^{17}+\cdots.$$

Thus we have a geometric interpretation of the coefficient b_{13}: a necessary and sufficient condition that the cubic ruled surface represented by equation (4) hyperosculate C at the undulation point O is that $b_{13}=a_{12}-a_{11}/2+a_9/8-a_7/16-15a/256$, which is equation (6) with $n=13$ since $a_5=-3a/2$. We shall henceforth assume that b_{13} does not have this value, so that the cubic ruled surface (4) does not hyperosculate C at O.

We now turn our attention to quartic cones intersecting C at O. If we write the most general homogeneous equation of the fourth degree in x, y, and z, this equation will represent a quartic cone with vertex at $O(1,0,0,0)$, the undulation point on the space curve C. By demanding that equations (2) for C satisfy this equation through terms in x^{15}, we have the equation of the 5-parameter family of cones having 16-point contact with C at O:

$$Fy^4+Ez^4+Gxz^3+Ay^3z+By^2z^2+Dyz^3=0, \tag{9}$$

where the capital letters represent homogeneous parameters. The following theorem is immediate:

THEOREM 2.6. *The tangent to a space curve C at an undulation point O is a triple line on any one of the 5-parameter family of quartic cones, with vertices at O, having 16-point contact with C at O.*

If we choose the unique one of these cones which has 21-point contact with C at O (that is, the osculating cone of C at O), its equation is found to be

$$16a^2y^4 - 16a^3xz^3 - (16aa_8 + 48aa_7 + 104aa_6 - 16a_6{}^2 - 75a^2)z^4 \\ - 4a(4a_6 - 3a)y^2z^2 - 2a(8a_7 + 20a_6 - 17a)yz^3 = 0.$$

Some of the properties of this cone are listed in the following:

THEOREM 2.7. *The tangent line of a space curve C at an undulation point O is a triple line on the osculating quartic cone of C at O. The intersection of the cone and the hyperosculating plane of C at O is the tangent of C at O counted four times. The intersection with the face $y = 0$ of the tetrahedron is the tangent counted three times and another line through the point O and in a plane through the edge $x = z = 0$. The intersection of the cone and the face $x = 0$ consists of four lines through the point O, each in a plane through the line $y = z = 0$.*

We next investigate the nature of some of the polar surfaces associated with this quartic cone. The polar plane of the vertex $(0, 0, 1, 0)$ of the tetrahedron is the face $y = 0$. The polar plane of the vertex $(0, 0, 0, 1)$ has the equation

$$8a^3x + a(8a_7 + 20a_6 - 17a)y + 2(16aa_8 + 48aa_7 + 104aa_6 \\ - 16a_6{}^2 - 75a^2)z = 0.$$

The polar quadric of the vertex $(0, 0, 1, 0)$ is degenerate, being composed of two planes given by the equation $24ay^2 - (4a_6 - 3a)z^2 = 0$. The polar cubic of the vertex $(0, 1, 0, 0)$ is the hyperosculating plane of C at O counted three times. The polar cubic of the vertex $(0, 0, 1, 0)$ is degenerate, being composed of three planes given by the equation

$$32ay^3 - 4(4a_6 - 3a)yz^2 - (8a_7 + 20a_6 - 17a)z^3 = 0.$$

The polar cubic of the vertex $(0, 0, 0, 1)$ is also degenerate, being composed of the hyperosculating plane of C at O and a quadric whose equation is

$$24a^3xz + 4a(4a_6 - 3a)y^2 + 3a(8a_7 + 20a_6 - 17a)yz \\ + 2(16aa_8 + 48aa_7 + 104aa_6 - 16a_6{}^2 - 75a^2)z^2 = 0.$$

3. Plane sections of the tangent developable. In the neighborhood of the undulation point O on the space curve C the equations of the tangent developable are

$$X = x + r, \qquad Y = y + ry', \qquad Z = z + rz', \tag{10}$$

where (x, y, z) represents the point on C at which the tangent is drawn, $(1, y', z')$ is a point on that tangent, primes denoting differentiation with respect to x, (X, Y, Z) is a variable point on the tangent, and r is a parameter. We shall first study the intersection of the tangent developable and the hyperosculating plane, $Z = 0$, of C at O. In order to do this, we set $Z = 0$ in the third of equations (10), substitute for z and z' their values from equations (2), and solve for r as a power series in x. Substitute this value for r in the first of equations (10), getting an expression for X in terms of x. Invert this series to get x as a power series in X. Then in the equation for Y in (10) substitute the series for r and the series for y and y' from (2), getting Y as a series in x. Now use the series for x in terms of X to get Y as a series in X. The result of this computation is

$$Y = (5^3/4^4)aX^4 - (3 \cdot 5^3/2 \cdot 4^4)aX^5 + \cdots, \quad Z = 0. \tag{11}$$

The following theorem follows from the form of equations (11):

THEOREM 3.1. *The section of the tangent developable in the neighborhood of an undulation point O on a space curve C made by the hyperosculating plane of the curve at the undulation point is a curve having an undulation at O, having as its tangent at O the undulational tangent of C at O, and having precisely 4-point contact with C at O.*

In a manner similar to that employed above, the intersection of the tangent developable with the face $X = 0$ of the tetrahedron of reference can be found. The equations of this section are

$$Z = AY^{5/4} + BY^{7/4} + \cdots, \qquad X = 0, \tag{12}$$

where

$$A = 4/\{3(-3a)^{1/4}\}, \qquad B = \{2 - (7a_6)/3a\}/\{3(-3a)^{3/4}\}.$$

From the nature of these equations, we draw the conclusion:

THEOREM 3.2. *The plane curve (12) has a quadruple point at O with the quadruple point tangents coinciding in the line $X = Y = 0$.*

Finally, let us investigate the section of the tangent developable made by the face $Y = 0$. The equations of this curve are

$$Z = AX^5 + BX^6 + CX^7 + \cdots, \qquad Y = 0,$$

where

$$A = (-4^4 a)/3^5, \qquad B = 6 \cdot 4^4/3^6, \qquad C = -4^4\{37a/2 - 4^2 a_6/3\}/3^7.$$

This section has a quintic point at O; the tangent at O is the undulational tangent, $Y = Z = 0$, of C at O; and the curve has precisely 4-point contact with C at O.

4. Projections of the curve. Let us take a point (α, β, γ) in space and project the curve C, represented by equations (2) in the neighborhood of the undulation point O on C, onto the faces of the tetrahedron of reference. If we designate a point on the curve as (x, y, z) and a point on the projection as (X, Y, Z), the equations of the projection are

$$(13) \quad X = x + (\alpha - x)t, \qquad Y = y + (\beta - y)t, \qquad Z = z + (\gamma - z)t,$$

where t is a parameter.

First we consider the projection of the curve onto the hyperosculating plane, $Z = 0$, of C at O. In this case $\gamma \neq 0$. To find the equations of this projection, set $Z = 0$ in (13), solve for t, put this value for t in the expression for X, at the same time substituting for z its power-series expansion from (2). This will give X as a series in x. Invert this series to get x as a series in X and then make this substitution for x in the expression for Y, having used the series for y from (2). The result of this computation furnishes the following equations for the projection:

$$(14) \quad Y = aX^4 + AX^5 + BX^6 + CX^7 + DX^8 + \cdots, \qquad Z = 0 \qquad (\gamma \neq 0),$$

where

$$A = -a(3/2 + \beta/\gamma), \qquad B = a_6 + (2a\beta)/\gamma,$$
$$C = a_7 - \beta(a_6 + 3a/4)/\gamma,$$
$$D = 4a^2\alpha/\gamma + a_8 - \beta(a_7 - a_6/2 + a/8)/\gamma.$$

From these equations the following theorem is seen to be true:

THEOREM 4.1. *The projection of a curve C in the neighborhood of an undulation point O from a point (α, β, γ) in space onto the hyperosculating plane of C at O is a curve having an undulation point at O, having the tangent of C at O as its tangent, and having precisely 5-point contact with C at O.*

It is evident that if (α, β, γ) is chosen in the face $X = 0$ but not on the line $X = Y = 0$, the results of the above theorem are not changed. However, if the point (α, β, γ) is in the plane $Y = 0$, but not on the line $X = Y = 0$, then equations (14) become

$$(15) \quad \begin{aligned} &Y = aX^4 - (3a/2)X^5 + a_6X^6 + a_7X^7 + (a_8 + 4a^2\alpha/\gamma)X^8 + \cdots, \\ &Z = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\gamma \neq 0), \end{aligned}$$

and it can be seen that this projection has, in addition to the properties stated above, the property that the cone represented by the series for Y in (15) has precisely 8-point contact with C at O.

If the point (α, β, γ) is chosen on the line $X = Y = 0$, but not at the point $(1, 0, 0, 0)$, the equations of the projection are

$$\begin{aligned} &Y = aX^4 - (3a/2)X^5 + a_6X^6 + a_7X^7 + a_8X^8 + (a_9 - 3a^2/\gamma)X^9 + \cdots, \\ &Z = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\gamma \neq 0). \end{aligned}$$

In addition to the properties stated in the theorem above, the cone represented by the series for Y has precisely 9-point contact with the curve at the undulation point.

Next we assume that the point (α, β, γ) is not in the face $X = 0$ (that is, $\alpha \neq 0$) and study the projection of the curve C from this point onto $X = 0$. Let us first assume that (α, β, γ) is not in the face $Y = 0$; the equations of the projection then are

$$Z = (\gamma/\beta)Y - (a\alpha^4\gamma/\beta^5)Y^4 + \cdots, \quad X = 0 \qquad (\alpha\beta \neq 0).$$

The next theorem follows immediately:

THEOREM 4.2. *If a space curve C, in the neighborhood of an undulation point O, be projected from a point (α, β, γ) not in either the face $X = 0$ or the face $Y = 0$, onto the face $X = 0$, then the projection is a curve having an undulation point or singularity of higher order at O. The tangent to the projection at O has the equations $X = 0$, $\beta Z - \gamma Y = 0$.*

If the point (α, β, γ) lies in the hyperosculating plane, $Z = 0$, but not on the line $Y = Z = 0$, the equations of the projection from $(\alpha, \beta, 0)$ onto the face $X = 0$ are:

$$Z = -\{a\alpha^5/\beta^5\}Y^5 - \{2a\alpha^5(2 + \alpha)/\beta^6\}Y^6 + \cdots, \qquad X = 0 \quad (\alpha\beta \neq 0).$$

This projection has a quintic point at the undulation point O of C and has as its tangent at the quintic point the edge $X = Z = 0$ of the tetrahedron.

If the point (α, β, γ) lies in the plane $Y = 0$ but not on the line

$Y=Z=0$, then the equations of the projection of C onto the face $X=0$ are

$$Y=AZ^4+BZ^5+CZ^6+\cdots,\qquad X=0 \qquad (\alpha\gamma\neq 0),$$

where

$$A=a\alpha^4/\gamma^4,\qquad B=(3a\alpha^4/\gamma^5)(1+\alpha/2),\qquad C=\alpha^4(\alpha^2a_6+6a+6a\alpha)/\gamma^6.$$

It can be seen that this projection has an undulation point at O with the edge $X=Y=0$ of the tetrahedron as the undulational tangent.

Now let us take (α,β,γ) on the line $Y=Z=0$, but not at the point $(1,0,0,0)$. The equations of this projection are

$$Z=a^{-1/4}Y^{5/4}-a^{-1/2}(1/\alpha+1/2)(Y^{3/2})/4+\cdots,\qquad X=0 \qquad (\alpha\neq 0).$$

This curve has a quadruple point at 0 with the quadruple point tangents coinciding in the edge $X=Y=0$ of the tetrahedron.

Lastly, let us take the projection of the curve C in the neighborhood of the undulation point O onto the face $Y=0$ from the point (α,β,γ) in space with $\beta\neq 0$. The equations of the projection are

$$(16)\quad Z=-(a\gamma/\beta)X^4+AX^5+BX^6+CX^7+DX^8+\cdots,\quad Y=0 \quad (\beta\neq 0),$$

where

$$A=a+3a\gamma/2\beta,\qquad B=-(2a+a_6\gamma/\beta),$$
$$C=a_6+3a/4-a_7\gamma/\beta-4a^2\alpha\gamma/\beta^2,$$
$$D=a_7-a_6/2+a/8-a_8\gamma/\beta+3a^2\gamma/\beta+27a^2\alpha\gamma/2\beta^2+5a^2\alpha/\beta.$$

The next theorem is a consequence of the form of these equations.

THEOREM 4.3. *The projection of C onto the plane $Y=0$, represented by equations* (16), *is a curve having an undulation point or singularity of higher order at O, the undulation point of C, and having precisely 4-point contact with C at O. The undulational tangent of the projection is the undulational tangent of C.*

It can easily be verified that if (α,β,γ) lies in the plane $X=0$, but not on either of the lines $X=Z=0$ or $X=Y=0$, none of the results of the above theorem is changed. On the other hand, if (α,β,γ) lies in the plane $Z=0$, that is, $\gamma=0$, but not on either of the lines $X=Z=0$ or $Y=Z=0$, then the projection represented by equations (16) has a quintic point at O with tangent $Y=Z=0$ and has 4-point contact with C at O. The cone represented by the equation for Z has precisely 8-point contact with C at O.

Finally, if (α, β, γ) lies on the line $X = Z = 0$ but not at the point $(1, 0, 0, 0)$, then the equations of the projection of C in the neighborhood of the undulation point O from $(0, \beta, 0)$ onto the plane $Y = 0$ are

$$Z = aX^5 - 2aX^6 + (a_6 + 3a/4)X^7 + (a_7 - a_6/2 + a/8)X^8 + (a_8 - a_7/2 - 3a/16 - 4a^2/\beta)X^9 + \cdots, \quad Y = 0 \qquad (\beta \neq 0).$$

This curve has a quintic point at O, the tangent at O being $Y = Z = 0$, the undulational tangent to C at O. The projection has precisely 4-point contact with C at O and the cone represented by the series for Z has precisely 9-point contact with C at O.

NEW JERSEY COLLEGE FOR WOMEN.

REFERENCES

1. G. Salmon, *A Treatise on the Analytic Geometry of Three Dimensions*, fifth ed., vol. 2, London, 1915.

2. E. J. Wilczynski, *Projective Differential Geometry of Curves and Ruled Surfaces*, Leipzig, 1906.

3. E. P. Lane, *A Treatise on Projective Differential Geometry*, Chicago, 1941.

4. A. T. Schafer, "Two singularities of space curves," *Duke Mathematical Journal*, vol. 11, no. 4 (1944), pp. 655-670.

PAIRS OF RECTILINEAR COMPLEXES.*

By V. G. Grove.

1. Introduction. It is our purpose in this paper to initiate a study of pairs of rectilinear complexes in one-to-one line correspondence. If corresponding lines of the pairs are skew we call the pair *skew*; if corresponding lines intersect we call the pair *non-skew*. We derive a canonical tetrahedron for each of these two cases. Fundamental to the study are pairs of pencils of linear complexes L_1, L_2 having first order contact with the given complexes Γ_1, Γ_2 along corresponding lines g_1, g_2.

Skew pairs may be classified into two types—to the first of these belong those pairs for which the congruence of lines common to certain complexes of the pencils L_1, L_2 has distinct directrices, and to the second those having coincident directrices.

Certain covariant figures associated with the pairs are found, namely quadrics and tetrahedral complexes, and general methods of finding covariant curves and surfaces are given. From these considerations many points, lines and planes, covariant to the pairs, may be found.

2. Analytical basis for skew pairs. Let the complexes Γ_1, Γ_2 under study be generated by lines g_1, g_2, and let x_1, x_2 be distinct points on g_1, and x_3, x_4 be distinct points on g_2. Let the homogeneous coordinates $(x_i^1, x_i^2, x_i^3, x_i^4)$ of these four points x_i be analytic functions of the independent variables u^1, u^2, u^3, and let the corresponding lines g_1, g_2 be given by the same values of u^1, u^2, u^3. The coordinates of the points x_i satisfy a system (3) of differential equations of the form

$$x_{i,\alpha} = a_{i\alpha}{}^j x_j, \qquad x_{i,\alpha} = \partial x_i / \partial u^\alpha. \tag{2.1}$$

repeated indices indicating summation over the ranges of those indices. The following indices, and ranges will be used consistently $i, j, k = 1, 2, 3, 4$; $\alpha, \beta, \rho, \sigma = 1, 2, 3$; $p, q = 1, 2$; $r, s = 3, 4$. The integrability conditions [1] of (2.1) are

$$a_{i\alpha}{}^j{}_{,\beta} + a_{i\alpha}{}^k a_{k\beta}{}^j = a_{i\beta}{}^j{}_{,\alpha} + a_{i\beta}{}^k a_{k\alpha}{}^j. \tag{2.2}$$

* Received June 28, 1947.

The complexes Γ_1, Γ_2 are unchanged by the tranformations

(2.3) $\qquad \bar{x}_p = \lambda_p{}^q x_q, \qquad \bar{x}_r = \mu_r{}^s x_s, \qquad |\lambda_p{}^q| \neq 0, \qquad |\mu_r{}^s| \neq 0,$
and

$$\bar{u}^\alpha = \theta^\alpha(u^1, u^2, u^3).$$

We shall use (2.3) to reduce (2.1) to a canonical form.

The general coordinates X^i of a point X may be related to its local coordinates ξ^i referred to the tetrahedron (x_1, x_2, x_3, x_4) by the expression

$$X^i = \xi^j x_j{}^i.$$

Let X_p be two points on a neighboring line g'_1 of g_1 of Γ_1 determined by the parameter values $u^\alpha + \Delta u^\alpha$, and X_r similarly be two points on g'_2 near g_2. The general coordinates of these four points are given by the power series expansions

$$X_i = x_i + x_{i,\rho}\Delta u^\rho + \cdots.$$

Using (2.1), we find that the power series expansions of the local coordinates $\xi_p{}^i$ of X_p and $\xi_r{}^i$ of X_r are respectively

(2.4) $\qquad \xi_p{}^i = \delta_p{}^i + a_{p\rho}{}^i \Delta u^\rho + \cdots, \xi_r{}^i = \delta_r{}^i + a_{r\rho}{}^i \Delta u^\rho + \cdots.$

Defining the homogeneous Plücker coordinates of the lines g'_1, g'_2 by the expressions

$$\omega_{12}{}^{ij} = \xi_1{}^i \xi_2{}^j - \xi_1{}^j \xi_2{}^i, \qquad \omega_{34}{}^{ij} = \xi_3{}^i \xi_4{}^j - \xi_3{}^j \xi_4{}^i,$$

we find from (2.4) that the power series expansions of these Plücker coordinates are

(2.5) $\qquad \omega_{12}{}^{ij} = \epsilon^{pq}\delta_p{}^i(\delta_q{}^j + a_{q\rho}{}^j \Delta u^\rho) + \cdots, \ \epsilon^{11} = \epsilon^{22} = \epsilon^{33} = \epsilon^{44} = 0$

$\qquad\qquad \omega_{34}{}^{ij} = \epsilon^{rs}\delta_r{}^i(\delta_s{}^j + a_{s\rho}{}^j \Delta u^\rho) + \cdots, \ \epsilon^{12} = \epsilon^{34} = -\epsilon^{21} = -\epsilon^{43} = 1.$

Imposing the conditions that the equations of the linear complexes

(2.6) $\qquad A_{ij}\omega^{ij} = 0, \qquad B_{ij}\omega^{ij} = 0, \qquad A_{ij} = -A_{ji}, \qquad B_{ij} = -B_{ji},$

be satisfied respectively by the first and second of (2.5) up to and including the first powers of Δu^α, we obtain *the equations of the pencils of linear complexes having four-line contact with* Γ_1, Γ_2 *along* g_1, g_2 *in the form* (2.6) *with coefficients given by*

$$(2.7) \qquad \epsilon^{pq}A_{pj}a_{qa}{}^{j}=0, \qquad \epsilon^{rs}B_{rj}a_{sa}{}^{j}=0, \qquad A_{12}=B_{34}=0,$$
$$A_{34}=k_1, \qquad B_{12}=k_2, \qquad k_1, k_2 \text{ arbitrary.}$$

Among the complexes of these pencils are those containing the lines g_2, g_1 respectively. They are given by (2.6) under the conditions (2.7) with $k_1=k_2=0$. We denote these by $L_1=0$, $L_2=0$ respectively. If we denote the well known invariants [2] [4] of $L_1=0$, $L_2=0$ by I_1, I_2 and their simultaneous invariant by J, that is

$$I_1=A_{12}A_{34}+A_{13}A_{42}+A_{14}A_{23}, \qquad I_2=B_{12}B_{34}+B_{13}B_{42}+B_{14}B_{23},$$
$$J=A_{12}B_{34}+A_{13}B_{42}+A_{14}B_{23}+A_{23}B_{14}+A_{42}B_{13}+A_{34}B_{12},$$

then the invariant I of the pencil $\lambda L_1+\mu L_2=0$ is given by

$$(2.8) \qquad I=I_1\lambda^2+J\lambda\mu+I_2\mu^2.$$

The directrices of the congruence $L_1=0$, $L_2=0$ are therefore distinct or coincident according as

$$T=J^2-4I_1I_2$$

does not or does vanish. We shall say the skew pair Γ_1, Γ_2 belongs to *the first type* or *the second type* according as $T\neq 0$ or $T=0$.

3. Skew pairs of the first type. In the null-system established by $L_1=0$ (or by the first of (2.6) under (2.7)) and by $L_2=0$ (or the second of (2.6)), the null-planes of the points $\bar{x}_p$, $\bar{x}_r$ defined by (2.3) have the respective equations

$$(3.1) \qquad \lambda^q{}_pA_{qr}\xi^r=0, \qquad \mu_r{}^sB_{ps}\xi^p=0.$$

The first of (3.1) for $p=1$ passes through $\bar{x}_4$, and for $p=2$ passes through $\bar{x}_3$, the second of (3.1) for $r=3$ passes through $\bar{x}_2$, and for $r=4$ through $\bar{x}_1$, if the respective equations are satisfied by $\lambda_q{}^p$, $\mu_s{}^r$:

$$(3.2) \quad \lambda_1{}^qA_{qr}\mu_4{}^r=0, \quad \lambda_2{}^qA_{qr}\mu_3{}^r=0, \quad \mu_3{}^rB_{pr}\lambda_2{}^p=0, \quad \mu_4{}^rB_{pr}\lambda_1{}^p=0.$$

Interpreting the first and third of (3.2) as two projectivities on a line corresponding "points" having coordinates $\lambda_1{}^1/\lambda_1{}^2$, $\mu_4{}^3/\mu_4{}^4$ and $\lambda_2{}^1/\lambda_2{}^2$, $\mu_3{}^3/\mu_3{}^4$, it is easy to show that the double points are distinct or coincident according as $T\neq 0$, or $T=0$. *Since for pairs of the first type* $T\neq 0$, *we may choose the point* x_1, x_2, x_3, x_4 *so that*

$$(3.3) \qquad A_{14}=A_{23}=B_{14}=B_{23}=0.$$

We may readily confirm the fact that if conditions (3.3) hold, and if I_1 (or $I_2 = 0$), then $T = 0$, contrary to hypothesis. It follows now from (2.7) and (3.3) that *there exist functions f, g such that*

$$(3.4) \qquad a_{2\alpha}{}^{3} = f a_{1\alpha}{}^{4}, \qquad a_{4\alpha}{}^{1} = g a_{3\alpha}{}^{2}.$$

The equations of the complexes $L_1 = 0$, $L_2 = 0$ now may be written in the respective forms

$$(3.5) \qquad L_1 = \omega^{13} - f\omega^{42} = 0, \qquad L_2 = \omega^{13} - g\omega^{42} = 0.$$

From (3.5) the special linear complexes of the pencil $\lambda L_1 + \mu L_2 = 0$ are the complexes $\omega^{13} = 0$, $\omega^{42} = 0$. *Hence the directrices of the congruence $L_1 = L_2 = 0$ are the lines (x_1x_3) and (x_2x_4).* The local tetrahedron of reference is therefore completely determined. *The complexes $L_1 = 0$, $L_2 = 0$ are in involution if $f + g = 0$.*

The vertices, faces and edges of the local tetrahedron being covariant to the pair Γ_1, Γ_2 one may find other covariant figures in many ways. We illustrate one method, which is essentially that of moving frames of reference of Cartan [1].

It may be verified that the derivatives of the local coordinates ξ^i of a point X are given by the formulas

$$\xi_{,\alpha}{}^{i} = - a_{j\alpha}{}^{i}\xi^{j}.$$

Under the conditions

$$(3.6) \qquad u^{\alpha} = u^{\alpha}(t), \qquad \lambda^{\alpha} = du^{\alpha}/dt,$$

the points x_i of the covariant tetrahedron of reference generate curves C_i, the lines g_1, g_2 generate ruled surfaces, and the faces $\xi^i = 0$ envelope developable surfaces S_i respectively. The tangents to the edges of regression of S_i are respectively

$$(3.7) \qquad \xi^{i} = a_{j\rho}{}^{i}\lambda^{\rho}\xi^{j} = 0.$$

Suppose the functions (3.6) satisfy the system of differential equations

$$(3.8) \qquad b_{\rho}{}^{p}\lambda^{\rho} = 0,$$

the matrix of whose coefficients is of rank two. Then the equations of the generators of S_i are

$$(3.9) \qquad \xi^{i} = (a_{j\alpha}{}^{i}\xi^{j}, b_{\alpha}{}^{1}, b_{\alpha}{}^{2}) = 0$$

wherein $(a_{j\alpha}{}^i, b_\alpha{}^1, b_\alpha{}^2)$ is the determinant of the matrix whose α-th row is $a_{j\alpha}{}^i, b_\alpha{}^1, b_\alpha{}^2$. In particular if $b_\alpha{}^1 = a_{3\alpha}{}^1$, $b_\alpha{}^2 = a_{4\alpha}{}^1$, *the line* g_2 *of* Γ_2 *is the tangent to the edge of regression of* S_i; again *if* $b_\alpha{}^1 = a_{2\alpha}{}^3 = fa_{1\alpha}{}^4$, $b_\alpha{}^2 = a_{4\alpha}{}^1 = ga_{3\alpha}{}^2$ *the tangents to the curves* C_i *generated by the points* x_i *intersect the respective edges* (x_2, x_3), (x_1, x_4), (x_1, x_4), (x_2, x_3) *for* $i = 1, 2, 3, 4$. *The edges of regression of the developables* S_i *are the respective curves* C_i *generated by* x_i. It will be shown in the last section that *the rank of the matrix of* (3.8) *in this latter case is one if and only if the four-line linear complexes of the complexes of lines* (x_1, x_3) *contain the lines* (x_1, x_4) *and* (x_2, x_3).

Returning to (2.6), under the conditions (3.4) we observe that the null-plane of the point $P_1(1, \lambda, 0, 0)$ on g_1 in the first of (2.6) intersects g_2 in the point $P_2(0, 0, \lambda f, -1)$. *The points* P_1 *on* g_1 *are projectively related to the points* P_2 *on* g_2, *the line* P_1P_2 *generating the quadric*

$$\xi^1\xi^3 + f\xi^2\xi^4 = 0. \tag{3.10}$$

By using the second of (2.6) a second covariant quadric,

$$\xi^1\xi^3 + g\xi^2\xi^4 = 0, \tag{3.11}$$

is found.

One may readily verify that *the polar planes of the points on* (x_1, x_3) $[(x_2, x_4)]$ *with respect to* (3.10) [(3.11)], *and the null-planes of those point with respect to* $L_1 = 0$ $(L_2 = 0)$ *are planes of an involution on the line* (x_2, x_4) $[(x_1, x_3)]$, *the double planes being the faces of the covariant tetrahedron through that line.*

Again, the tangent planes at points on g_1 to the quadrics (3.10), (3.11) intersects g_2 in projective ranges of points. These points determine respectively with x_1, x_2 projectively related pencils of lines. The tetrahedral complex [2] determined by these pencils has the equation.

$$g\omega^{14}\omega^{23} + f\omega^{13}\omega^{42} = 0. \tag{3.12}$$

One of the cross ratios of the points of intersection of the lines of this tetrahedral complex being f/g, we note that *these points are harmonic if and only if* $L_1 = 0$ *and* $L_2 = 0$ *are in involution, that is* $f + g = 0$.

The ruled surface R_1 *generated by* g_1, *under the conditions* (3.6), *is a developable if and only if*

$$\epsilon^{pq}a_{p\rho}{}^3a_{q\sigma}{}^4du^\rho du^\sigma = 0, \tag{3.13}$$

and the ruled surface R_2 generated by g_2 is a developable if and only if

$$\epsilon^{rs}a_{r\rho}{}^1a_{s\sigma}du^\rho du^\sigma = 0. \tag{3.14}$$

It follows from (3.13) and (3.14) that ordinarily *there exist four curves* C_1 *along which* x_1 *must move so that the ruled surfaces* R_1, R_2 *generated by* g_1, g_2 *shall both be developables.* A partial study of skew pairs of the first type may be made by a study of the simultaneous quadratics (3.13), (3.14).

4. Skew pairs of the second type. In a manner similar to that used in the previous section, we may choose the line (x_1, x_3) as the doubly counted directrix of the congruence $L_1 = 0$, $L_2 = 0$. This is equivalent to choosing the coordinate system so that $A_{13} = B_{13} = 0$, as may readily be verified from (2.7) and the condition $T = 0$. An examination of I_1, I_2 and $T = 0$ shows the existence of a function g such that

$$(4.1) \qquad B_{23} = gA_{23}, \qquad B_{14} = gA_{14}.$$

The pencil $\lambda L_1 + \mu L_2 = 0$ has a doubly counted special complex given by $\lambda/\mu = -g$, the equation of this complex being $\omega^{42} = 0$ verifying that the line (x_1, x_3) is the directrix of $L_1 = L_2 = 0$, provided of course that $B_{42} \neq gA_{42}$. *If* $B_{42} = gA_{42}$, *the complexes* $L_1 = 0$, $L_2 = 0$ *are identical. Since a single linear complex has no absolute invariant, it is clear that no covariant tetrahedron can be found for* Γ_1, Γ_2 *using such linear complexes.*

Under the conditions (4.1), the equations of the complexes $L_1 = 0$, $L_2 = 0$ asume the form

$$(4.2) \qquad \begin{aligned} L_1 &= A_{14}\omega^{14} + A_{23}\omega^{23} + A_{42}\omega^{42} = 0, \\ L_2 &= g[A_{14}\omega^{14} + A_{23}\omega^{23}] + B_{42}\omega^{42} = 0. \end{aligned}$$

The form of (4.2) is preserved under all transformations of the form

$$(4.3) \qquad \bar{x}_1 = x_1, \quad \bar{x}_2 = \lambda_2{}^p x_p, \quad \bar{x}_3 = x_3, \quad \bar{x}_4 = \mu_4{}^r x_r, \quad \lambda_2{}^2\mu_4{}^4 \neq 0.$$

From the first of (4.2) we observe that the null-plane of x_2 in $L_1 = 0$ is the plane (x_1, x_2, x_4) if and only if $A_{42} = 0$. An examination of the transform $\bar{A}_{42}$ of A_{42} under (4.3) shows that *for a given point* $\bar{x}_2$ *there exists a unique point* $\bar{x}_4$ *for which* $\bar{A}_{42} = 0$. The conditions $A_{42} = 0$, $A_{13} = 0$, in view of (2.7), and $I_1 \neq 0$ imply the existence of a function f such that

$$a_{2a}{}^4 = f a_{1a}{}^3.$$

The equations $L_1 = 0$, $L_2 = 0$ now assume the form

$$(4.4) \qquad \begin{aligned} L_1 &= \omega^{14} + f\omega^{23} = 0 \\ L_2 &= g(\omega^{14} + f\omega^{23}) + h\omega^{42} = 0, \qquad h = B_{42}/A_{14}. \end{aligned}$$

The form of (4.4) is preserved under all transformations of the form (4.3) for which

(4.5) $$\lambda_2^1/\lambda_2^2 + f\mu_4^3/\mu_4^4 = 0.$$

The projectivity (4.5) *between the points* $\bar{x}_2$ *on* g_1 *and* $\bar{x}_4$ *on* g_2 *generates the covariant quadric*

(4.6) $$\xi^1\xi^4 + f\xi^2\xi^3 = 0.$$

It is easy to show that *the envelope of the null-planes in* $L_1 = 0$ *of points on* $(\bar{x}_2, \bar{x}_4)$ *is the quadric* (4.6); *and the envelope of the null-planes in* $L_2 = 0$ *of the points on* $(\bar{x}_2, \bar{x}_4)$ *is the quadric*

$$g(\xi^1\xi^4 + f\xi^2\xi^3) - 2h\xi^2\xi^4 = 0.$$

In a manner similar to that used in Section 2 *the equation of the pencil of four-line complexes of the complex of lines* (x_2, x_3) *is*

(4.7) $$C_{ij}\omega^{ij} = 0, \qquad C_{23} = 0, \qquad C_{14} = k\ \text{(arbitrary)}$$

the remaining coefficients being determined by the equations

(4.8) $$C_{2j}a_{3\rho}{}^j - C_{3j}a_{2\rho}{}^j = 0.$$

We observe that *the covariant line* (x_1, x_3) *belongs to every complex of the pencil* (4.7) *if and only if* $C_{13} = 0$. Under the transformation (4.3), the transform $\bar{C}_{13}$ of C_{13} is given by the expression

$$\begin{aligned}\bar{C}_{13} = -\lambda_2{}^2\{&(a_{3\alpha}{}^4, a_{3\alpha}{}^2, a_{1\alpha}{}^4)(\lambda_2{}^1/\lambda_2{}^2)^2 \\ &+ [(a_{3\alpha}{}^4, a_{3\alpha}{}^2, a_{2\alpha}{}^4) - (a_{3\alpha}{}^4, a_{3\alpha}{}^1, a_{1\alpha}{}^4)]\lambda_2{}^1/\lambda_2{}^2 - C_{13}\}/\mu_4{}^4.\end{aligned}$$

There exist therefore two points $\bar{x}_2$ *on* g_1 *each of which determines with the covariant point* x_3 *a complex whose pencil of four-line complexes contains the covariant line* (x_1, x_3). We now choose the point x_2 as the harmonic conjugate of x_1 with respect to these points. That is we may choose x_2 so that

(4.9) $$(a_{3\alpha}{}^1, a_{3\alpha}{}^4, a_{1\alpha}{}^4) = (a_{3\alpha}{}^2, a_{3\alpha}{}^4, a_{2\alpha}{}^4), \ (a_{3\alpha}{}^2, a_{3\alpha}{}^4, a_{1\alpha}{}^4) \neq 0.$$

The point x_4 is determined by the projectivity (4.5). This coordinate system may be described as follows: *the tetrahedron* x_1, x_2, x_3, x_4 *is such that the lines* (x_1, x_2), (x_3, x_4) *are the lines* g_1, g_2 *of* Γ_1, Γ_2, *the edge* (x_1, x_3) *is the doubly counted directrix of the congruence* $L_1 = L_2 = 0$; *the vertex* x_2 *is the harmonic conjugate of* x_1 *with respect to the two points each of which determines with* x_3 *a complex whose four-line complexes each contain this directrix*;

the point x_4 is the corresponding point of x_2 in the projectivity which preserves the form of $L_1 = 0$, $L_2 = 0$. We give a geometrical interpretation of the inequality in (4.9) in the next section.

In a manner similar to that of the previous section, covariant curves, lines and points for pairs of this type may be found. We content ourselves with the remark that *on every surface through x_1 there is a curve C_1 whose tangent at x_1 intersects the edge (x_2, x_3) of the covariant tetrahedron. The tangent to the corresponding curve C_2 at x_2 intersects the directrix (x_1, x_3) of $L_1 = L_2 = 0$.*

5. Non-skew pairs. In this section we derive a canonical tetrahedron for non-skew pairs of complexes. Let the corresponding lines g_1, g_2 of Γ_1, Γ_2 be the lines (x_1, x_3), (x_1, x_4) respectively in the notation of Section **2**. These complexes are unchanged by the transformation

$$(5.1) \qquad \bar{x}_1 = x_1, \quad \bar{x}_2 = \lambda^j x_j, \quad \bar{x}_3 = l x_1 + m x_4, \quad \bar{x}_4 = \lambda x_1 + \mu x_4.$$

The four line complexes of Γ_1, Γ_2 along g_1, g_2 are found to be respectively

$$(5.2) \qquad P_{ij}\omega^{ij} = 0; \quad Q_{ij}\omega^{ij} = 0, \quad P_{13} = Q_{14} = 0, \quad P_{42} = k_1, \quad Q_{23} = k_2,$$

k_1, k_2 arbitrary, the remaining coefficients being determined by the equations

$$(5.3) \qquad P_{1j}a_{3a} - P_{3j}a_{1a} = 0, \qquad Q_{1j}a_{4a}{}^j - Q_{4j}a_{1a}{}^j = 0.$$

The first of (5.2) contains *the lines* (x_1, x_4). (x_2, x_3) *if* $P_{14} = P_{23} = 0$. These conditions are equivalent to demanding that the rank of (3.8) with $b_a{}^1 = a_{3a}{}^1$, $b_a{}^2 = a_{4a}{}^1$ be one.

We observe that *the null-plane of x_3 in the first of* (5.2) *passes through x_4 and the null-plane of the second passes through x_3 respectively if and only if*

$$(5.4) \qquad P_{34} = Q_{34} = 0.$$

The transformation (5.1) with

$$P_{14} l + P_{34} m = 0, \qquad Q_{13}\lambda - Q_{34}\mu = 0$$

makes $\bar{P}_{34}$, $\bar{Q}_{34}$, the transforms of P_{34}, Q_{34} under (5.1), vanish. We assume that this transformation has been effected. It remains to choose the edge (x_1, x_2) and the vertex x_2 of the tetrahedron.

To this end we choose the line (x_1, x_2) to lie in both complexes (5.2). This is equivalent to choosing λ^2, λ^3, λ^4 in (5.1) so that $P_{12} = Q_{12} = 0$.

This choice is unique; we assume that this transformation has been effected. Assuming that the complexes (5. 2) are not special, that is $P_{14}\cdot P_{23}\neq 0$, $Q_{13}\cdot Q_{42}\neq 0$, it follows from our assumptions concerning the choice of x_3, x_4 and the line (x_1, x_2), namely

$$P_{12}=P_{34}=Q_{12}=Q_{34}=0,$$

that there exist non-vanishing functions f and g such that

$$a_{3a}{}^4=fa_{1a}{}^2,\qquad a_{4a}{}^3=-ga_{1a}{}^2. \tag{5. 5}$$

The four line complexes of Γ_1, Γ_2 along g_1, g_2 assume the respective simple forms

$$\omega^{14}-f\omega^{23}+k_1\omega^{42}=0,\qquad \omega^{13}-g\omega^{42}+k_2\omega^{23}=0.$$

The conditions (5. 5) are invariant under the transformation

$$\bar{x}_1=x_1,\qquad \bar{x}_2=\lambda^p x_p,\qquad \bar{x}_3=x_3,\qquad \bar{x}_4=x_4,\qquad \lambda^2\neq 0.$$

The equation of the pencil of four-line linear complexes of the complex of lines (x_1, x_2) is

$$A_{ij}\omega^{ij}=0, \tag{5. 6}$$

A_{ij} being defined by the first of (2. 7) and $A_{12}=0$, $A_{34}=k$. This complex contains the line (x_3, x_4) if $k=0$, and the null-planes of x_3, x_4 are the respective planes

$$A_{p3}\xi^p=0,\qquad A_{p4}\xi^p=0. \tag{5. 7}$$

The point x_2 may be uniquely chosen as the harmonic conjugate of x_1 with respect to the two points of intersection of the planes (5. 7) with the line (x_1, x_2). That is we may choose x_2 so that

$$A_{23}A_{14}-A_{42}A_{13}=0.$$

The equation of the pencil (5. 6) *then assumes the form*

$$A_{13}[\omega^{13}+h\omega^{23}]+A_{14}[\omega^{14}+h\omega^{42}]+k\omega_{34}=0.$$

A covariant tetrahedron for non-skew pairs of complexes has therefore been characterized geometrically. It may be described as follows: *the edges*

(x_1, x_3), (x_1, x_4) are the corresponding line g_1, g_2 of the pair; the edge (x_1, x_2) belongs to all complexes of the pencils of four-line linear complexes of Γ_1, Γ_2 along g_1, g_2; the null-planes of x_4 and x_3 in these respective complexes pass through x_3 and x_4 respectively; the point x_2 is the harmonic conjugate of x_1 with respect to the two points of intersection with (x_1, x_2) of the null-planes of x_3, x_4 in the four-line linear complex of the complex of lines (x_1, x_2) which contains (x_3, x_4).

Returning to the first of (5.2) with coefficients given by the first of (5.3), we find that the equation of the null-plane of the point X defined by

$$X = \lambda^1 x_1 + \lambda^3 x_3$$

is

$$\lambda^1(P_{12}\xi^2 + P_{14}\xi^4) + \lambda^3(P_{23}\xi^2 + P_{43}\xi^4) = 0.$$

By imposing the conditions that this plane coincides first with the covariant plane (x_1, x_3, x_4) and then the plane (x_1, x_2, x_3) the covariant points

$$(5.8) \qquad X_1 = P_{43}x_1 + P_{14}x_3, \qquad X_2 = P_{32}x_1 + P_{12}x_3$$

are given geometrical characterizations. The vanishing of P_{14} and P_{23} may therefore be interpreted. But from (5.3) we find that

$$P_{14} = (a_{3a}{}^2, a_{1a}{}^4, a_{1a}{}^2), P_{23} = (a_{3a}{}^2, a_{3a}{}^4, a_{1a}{}^4).$$

A geometrical interpretation of the inequality in (4.9) is therefore known.

In a manner similar to that used in Section **4**, and in view of (5.5) we may show *that on every surface through x_1 there exists a curve C_1 such that its tangent at x_1 intersects the line (x_3, x_4). The tangents to the corresponding curves C_3, C_4 at x_3, x_4 intersect the line (x_1, x_2). Unless the rank of the matrix of the coefficients of the equations*

$$(5.9) \qquad a_{1\rho}{}^2\lambda^\rho = 0, \qquad a_{2\rho}{}^1\lambda^\rho = 0$$

is less than two, there exist unique curves C_1, C_2, C_3, C_4 whose tangents at x_1, x_2, x_3, x_4 intersect the respective lines (x_3, x_4), (x_3, x_4), (x_1, x_2), (x_1, x_2). If the rank of (5.9) is one, there exists a function h such that

$$a_{2a}{}^1 = h a_{1a}{}^2.$$

From (4.7) and (4.8) *these conditions are equivalent to demanding that the four-line complexes of the complex of lines (x_2, x_3) be non-singular and contain the lines (x_1, x_2), (x_3, x_4).*

Under the conditions (5.9), the tangents to the edges of regression of the developable surface S_i enveloped by the planes $\xi^i = 0$ are

$$\xi^i = (a_{ja}{}^i \xi^j, a_{2a}{}^1, a_{1a}{}^2) = 0$$

These covariant lines determine in an obvious manner other covariant points and lines.

MICHIGAN STATE COLLEGE.

REFERENCES.

1. É. Cartan, *Leçons sur la théorie des espaces à connexion projective*, Paris, Gauthier-Villars (1937).
2. C. M. Jessop, *A treatise on the line complex*, Cambridge University Press (1903).
3. E. P. Lane, "The projective differential geometry of systems of linear homogeneous differential equations of the first order," *Transactions of the American Mathematical Society*, vol. 30 (1928), pp. 786-796.
4. E. J. Wilzynski, *Projective differential geometry of curves and ruled surfaces*, Leipzig, Teubner (1906).

ON THE REPRESENTATION OF PROJECTIVE ALGEBRAS.*

By J. C. C. McKinsey.

C. J. Everett and S. Ulam [1] have defined projective algebras and have raised the question whether every projective algebra is isomorphic to a certain kind of class of binary relations. They answer this question affirmatively for complete atomic projective algebras. In the present paper I shall show that the answer is affirmative also for the general case, by proving that every projective algebra is isomorphic to a subalgebra of a complete atomic projective algebra.

The method of proof is by means of the Boolean ideals which M. H. Stone used in establishing a representation theorem for Boolean algebras.[2] I shall assume without proof such well-known results [3] as that every ideal is contained in a prime ideal; and that, if one prime ideal is contained in another, then the two are identical.

I shall refer to Everett and Ulam's postulates and elementary theorems by the same symbols they use ("P1", $\cdots$, "P7" for the postulates, and "C1", $\cdots$, "C31" for the theorems), but without quoting them *in extenso*.

I shall use Everett and Ulam's notations for Boolean algebras: thus I use "$\wedge$" for logical product, "$\vee$" for logical sum, an accent for negation, "i" for the Boolean unit element, "0" for the Boolean zero element, "$\leqq$" for inclusion, and "$<$" for proper inclusion. I also use their notation for x-projection, y-projection, and $\square$-product.

I shall follow Everett and Ulam in using lower case Latin letters to designate elements of an arbitrary projective algebra. I shall use lower case Greek letters to indicate sets of elements of an arbitrary projective algebra—and thus, in particular, for ideals. I shall use Latin capitals to stand for classes of sets of elements of arbitrary projective algebras—thus, in particular, for classes of ideals.

I shall use the usual notation of set theory for set-theoretical operations

* Received February 27, 1947.

[1] C. J. Everett and S. Ulam, "Projective algebra I," *American Journal of Mathematics*, vol. 68 (1946), pp. 77-88.

[2] M. H. Stone, "The theory of representations for Boolean algebras," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 37-111.

[3] See M. H. Stone, *loc. cit.*

and relations—thus, " $\bigcup$ " for union, " $\bigcap$ " for intersection, " — " for difference, " $\subseteq$ " for inclusion, " $\subset$ " for proper inclusion, and " ε " for membership. I shall write " $\{a\}$ " to mean the set whose only member is a, and " Λ " for the empty set.

Definition 1. By an *ideal* in a Boolean algebra, I shall mean a non-empty set α of elements such that:

(i) if $a \,\varepsilon\, \alpha$ and $b \,\varepsilon\, \alpha$, then $a \wedge b \,\varepsilon\, \alpha$;

(ii) if $a \,\varepsilon\, \alpha$, and $a \leqq b$, then $b \,\varepsilon\, \alpha$;

(iii) $0 \,\bar{\varepsilon}\, \alpha$.

An ideal α is called a *prime* ideal, if it satisfies the further condition:

(iv) for every a, either $a \,\varepsilon\, \alpha$ or $a' \,\varepsilon\, \alpha$.

Definition 2. If α is any prime ideal, then by α_x we shall mean the set of all elements u such that, for some v in α, $v_x \leqq u$. By α_y we shall mean the set of all elements u such that, for some v in α, $v_y \leqq u$.

LEMMA 1. *If α is a prime ideal, and if $a \leqq i_x$, then either $a \,\varepsilon\, \alpha_x$ or $(a' \wedge i_x) \,\varepsilon\, \alpha_x$.*

Proof. Of the two elements $a \,\Box\, i_y$ and $(a \,\Box\, i_y)'$, one belongs to α. Hence, by C21, either $(a \,\Box\, i_y) \,\varepsilon\, \alpha$ or $[(a' \wedge i_x) \,\Box\, i_y] \,\varepsilon\, \alpha$. Hence, by Definition 2, either $(a \,\Box\, i_y)_x \,\varepsilon\, \alpha_x$ or $[(a' \wedge i_x) \,\Box\, i_y]_x \,\varepsilon\, \alpha_x$. Hence, by P5, either $a \,\varepsilon\, \alpha_x$ or $(a' \wedge i_x) \,\varepsilon\, \alpha_x$.

LEMMA 2. *If α is an ideal, then α_x is an ideal.*

Proof. Condition (i) of Definition 1 follows from C4. Condition (ii) is obvious. Condition (iii) follows from P3.

LEMMA 3. *If α is a prime ideal, then α_x is a prime ideal.*

Proof. From Lemma 2, we see that it suffices to show that condition (iv) of Definition 1 is satisfied. Let a be any element of the projective algebra. Since $(a \wedge i_x) \leqq i_x$, we see from Lemma 1 that either $(a \wedge i_x) \,\varepsilon\, \alpha_x$ or $[(a \wedge i_x)' \wedge i_x] \,\varepsilon\, \alpha_x$. Since $[(a \wedge i_x)' \wedge i_x] = a' \wedge i_x$, we conclude that either $(a \wedge i_x) \,\varepsilon\, \alpha_x$ or $(a' \wedge i_x) \,\varepsilon\, \alpha_x$. Since $(a \wedge i_x) \leqq a$ and $(a' \wedge i_x) \leqq a'$, our lemma now follows from condition (ii) of Definition 1.

LEMMA 4. *If α is a prime ideal, then α_y is a prime ideal.*

Proof. Similar to the proof of Lemma 3.

Definition 3. If A is a class of prime ideals, then by A_x we shall mean the class of all prime ideals α_x, where $\alpha \varepsilon A$. By A_y we shall mean the class of all prime ideals α_y, where $\alpha \varepsilon A$. We shall denote the class of all prime ideals by I; thus I_x is the class of all prime ideals α such that, for some β, $\alpha = \beta_x$.

LEMMA 5. *If A and B are classes of prime ideals, then*

$$(A \cup B)_x = A_x \cup B_x \text{ and } (A \cup B)_y = A_y \cup B_y.$$

Proof. From Definition 3.

LEMMA 6. *If A is a class of prime ideals, then $A_x = \Lambda$ if and only if $A = \Lambda$, and $A_y = \Lambda$ if and only if $A = \Lambda$.*

Proof. By Definition 3 and Lemmas 3 and 4.

LEMMA 7. *If a is any element of a projective algebra, then $(a \wedge i_x) \leqq a_x$.*

Proof. Since $(a \wedge i_x) \leqq a$, we see by C1 that $(a \wedge i_x)_x \leqq a_x$. Since $(a \wedge i_x) \leqq i_x$, we see by C8 that $(a \wedge i_x)_x = (a \wedge i_x)$. Hence the lemma.

LEMMA 8. *For any prime ideal α, the following conditions are equivalent.*

(i) $\alpha = \alpha_x$;

(ii) $\alpha \varepsilon I_x$;

(iii) $i_x \varepsilon \alpha$.

Proof. It is obvious that (i) implies (ii).

To see that (ii) implies (iii), let $\alpha \varepsilon I_x$. Then there exists a prime ideal β such that $\alpha = \beta_x$. Since $i \varepsilon \beta$, we conclude that $i_x \varepsilon \alpha$.

To see that (iii) implies (i), suppose that $i_x \varepsilon \alpha$. In order to show that $\alpha = \alpha_x$ it suffices (since α and α_x are prime ideals) to show that $\alpha_x \subseteq \alpha$. Let a be an arbitrary element of α_x. Then, for some b in α, $b_x \leqq a$. Since $b \varepsilon \alpha$, and, by hypothesis, $i_x \varepsilon \alpha$, we have $(b \wedge i_x) \varepsilon \alpha$, and hence, by Lemma 7, $b_x \varepsilon \alpha$. Hence, since $b_x \leqq a$, we have $a \varepsilon \alpha$, as was to be shown.

LEMMA 9. *For any prime ideal α, the following conditions are equivalent*

(i) $\alpha = \alpha_y$;

(ii) $\alpha \varepsilon I_y$;

(iii) $i_y \varepsilon \alpha$.

Proof. Similar to the proof of Lemma 8.

LEMMA 10. *If α is any prime ideal, then $\alpha_{xx} = \alpha_x$ and $\alpha_{yy} = \alpha_y$.*

Proof. By Lemmas 8 and 9, since $\alpha_x \varepsilon I_x$ and $\alpha_y \varepsilon I_y$.

LEMMA 11. *If A is any class of prime ideals, then $A_{xx} = A_x$ and $A_{yy} = A_y$.*

Proof. By Lemma 10 and Definition 3.

Definition 4. By π_0 we shall mean the set of all elements u of the projective algebra such that $p_0 \leqq u$, where p_0 is the atom mentioned in P2 and P6. By P_0 we shall mean the class whose only member is π_0.

Remark. It is readily seen that π_0 is a prime ideal.

LEMMA 12. *If α is any prime ideal, then $\alpha_{xy} = \alpha_{yx} = \pi_0$.*

Proof. To show that $\alpha_{xy} = \pi_0$, it suffices to show that $p_0 \varepsilon \alpha_{xy}$. But since $i \varepsilon \alpha$, we have $i_x \varepsilon \alpha_x$, and hence $i_{xy} \leqq \alpha_{xy}$; therefore, using P2, $p_0 \varepsilon \alpha_{xy}$, as was to be shown. The proof that $\alpha_{yx} = \pi_0$ is similar.

LEMMA 13. $I_{xy} = I_{yx} = P_0$.

Proof. By Definition 3 and Lemma 12.

Definition 5. If α and β are prime ideals such that $\alpha \varepsilon I_x$ and $\beta \varepsilon I_y$, then we denote by $\alpha \boxtimes \beta$ the set of all elements u such that, for some v in α and w in β, $u = (v \wedge i_x) \square (w \wedge i_y)$. We denote by $\alpha \boxdot \beta$ the set of all elements u such that, for some v in $\alpha \boxtimes \beta$, $v \leqq u$.

LEMMA 14. *If α and β are prime ideals such that $\alpha \varepsilon I_x$ and $\beta \varepsilon I_y$ then $\alpha \boxdot \beta$ is an ideal.*

Proof. To show that condition (i) of Definition 1 is satisfied, suppose

that $u_1 \varepsilon (\alpha \boxdot \beta)$ and $u_2 \varepsilon (\alpha \boxdot \beta)$. Then there are elements a_1 and a_2 of α, and elements b_1 and b_2 of β, such that

$$[(a_1 \wedge i_x) \square (b_1 \wedge i_y)] \leqq u_1$$
$$[(a_2 \wedge i_x) \square (b_2 \wedge i_y)] \leqq u_2;$$

from this we conclude immediately that

(1) $\{[(a_1 \wedge i_x) \square (b_1 \wedge i_y)] \wedge [(a_2 \wedge i_x) \square (b_2 \wedge i_y)]\} \leqq (u_1 \wedge u_2)$.

From C11, since $a_1 \wedge i_x \leqq i_x$, $a_2 \wedge i_x \leqq i_x$, $b_1 \wedge i_y \leqq i_y$, and $b_2 \wedge i_y \leqq i_y$, we see that

(2) $\{[(a_1 \wedge i_x) \wedge (a_2 \wedge i_x)] \square [(b_1 \wedge i_y) \wedge (b_2 \wedge i_y)]\}$
$= \{[(a_1 \wedge i_x) \square (b_1 \wedge i_y)] \wedge [(a_2 \wedge i_x) \square (b_2 \wedge i_y)]\}$.

It is clear, since we are dealing with a Boolean algebra, that

(3) $$[(a_1 \wedge i_x) \wedge (a_2 \wedge i_x)] = [a_1 \wedge a_2 \wedge i_x]$$
and
(4) $$[(b_1 \wedge i_y) \wedge (b_2 \wedge i_y)] = [b_1 \wedge b_2 \wedge i_y].$$

From (1) we conclude, by means of (2), (3), and (4) that

(5) $$[(a_1 \wedge a_2 \wedge i_x) \square (b_1 \wedge b_2 \wedge i_y)] \leqq (u_1 \wedge u_2).$$

Since $(a_1 \wedge a_2) \varepsilon \alpha$ and $(b_1 \wedge b_2) \varepsilon \beta$, we see that

$$[(a_1 \wedge a_2 \wedge i_x) \square (b_1 \wedge b_2 \wedge i_y)]$$

is in $\alpha \boxtimes \beta$. Hence, from (5), $u_1 \wedge u_2$ is in $\alpha \boxdot \beta$, as was to be shown.

It is obvious that condition (ii) is satisfied.

To see that condition (iii) is satisfied let us consider the implications of the assumption that $0 \varepsilon \alpha \boxdot \beta$. Then by Definition 5 we conclude that $0 \varepsilon \alpha \boxtimes \beta$, and hence that there is an element a of α and an element b of β such that $(a \wedge i_x) \square (b \wedge i_y) = 0$. By C7 we conclude that either $a \wedge i_x = 0$ or $b \wedge i_y = 0$. By Lemmas 8 and 9, $i_x \varepsilon \alpha$ and $i_y \varepsilon \beta$; hence since α and β are ideals, $(a \wedge i_x) \varepsilon \alpha$ and $(b \wedge i_y) \varepsilon \beta$. Thus we see that either $0 \varepsilon \alpha$ or $0 \varepsilon \beta$, which is absurd in view of the fact that α and β are ideals.

Definition 6. If α and β are prime ideals such that $\alpha \varepsilon I_x$ and $\beta \varepsilon I_y$, then by $\alpha \square \beta$ we shall mean the class of all prime ideals γ such that $(\alpha \boxdot \beta) \subseteq \gamma$. If A and B are classes of prime ideals such that $A \subseteq I_x$ and

$B \subseteq I_y$ then by $A \square B$ we shall mean the union of all classes $\alpha \square \beta$ of prime ideals, where $\alpha \varepsilon A$ and $\beta \varepsilon B$.

Lemma 15. *If A and B are non-empty classes of prime ideals, such that $A \subseteq I_x$ and $B \subseteq I_y$, then $A \square B$ is not empty.*

Proof. Let $\alpha \varepsilon A$ and $\beta \varepsilon B$. By Lemma 14, $\alpha \boxdot \beta$ is an ideal. Since every ideal is contained in a prime ideal, we see from Definition 6 that $\alpha \square \beta$ is not empty. Since $\alpha \square \beta$ is contained in $A \square B$, we conclude, finally, that $A \square B$ is not empty.

Lemma 16. *If A and B are classes of prime ideals such that $A \subseteq I_x$ and $B \subseteq I_x$, then $(A \cup B) \square I_y = (A \square I_y) \cup (B \square I_y)$. If A and B are classes of prime ideals such that $A \subseteq I_y$ and $B \subseteq I_y$ then $I_x \square (A \cup B) = (I_x \square A) \cup (I_x \square B)$.*

Proof. By the second part of Definition 6.

Lemma 17. *If $\alpha \varepsilon I_x$, then $(\alpha \boxdot \pi_0) = \alpha$.*

Proof. Since α is a prime ideal, and $\alpha \boxdot \pi_0$ is an ideal, it suffices to show that $\alpha \subseteq \alpha \boxdot \pi_0$. Suppose, then, that a is an arbitrary element of α. Since, by Definition 4, $p_0 \varepsilon \pi_0$, we see by Definition 5 that $[(a \wedge i_x) \square (p_0 \wedge i_y)] \varepsilon (\alpha \boxdot \pi_0)$. By C6 we have $p_0 \wedge i_y = p_0$, and by C16 we see that $(a \wedge i_x) \square p_0 = (a \wedge i_x)$. Thus we conclude that $(a \wedge i_x) \varepsilon (\alpha \boxdot \pi_0)$. Since $a \wedge i_x \leqq a$ we therefore have $a \varepsilon (\alpha \boxdot \pi_0)$, as was to be shown.

Lemma 18. *If $\alpha \varepsilon I_x$, then $\alpha \square \pi_0 = \{\alpha\}$.*

Proof. It is clear, by Definition 6, that $\alpha \varepsilon (\alpha \square \pi_0)$ since, by Lemma 17, $\alpha = (\alpha \boxdot \pi_0)$. Now let β be any prime ideal belonging to $\alpha \square \pi_0$. Then $\alpha \boxdot \pi_0 \subseteq \beta$, and hence, by Lemma 17, $\alpha \subseteq \beta$. Since α and β are prime ideals, we therefore conclude that $\beta = \alpha$, as was to be shown.

Lemma 19. *If $\alpha \varepsilon I_y$, then $\pi_0 \square \alpha = \{\alpha\}$.*

Proof. Similar to the proof of Lemma 18.

Lemma 20. *$I_x \square P_0 = I_x$ and $P_0 \square I_y = I_y$.*

Proof. By Lemmas 18 and 19.

Lemma 21. *Let $\alpha \varepsilon I_x$ and $\beta \varepsilon I_y$, and suppose that $\gamma \varepsilon (\alpha \square \beta)$; then $\gamma_x = \alpha$ and $\gamma_y = \beta$.*

Proof. To show that $\gamma_x = \alpha$, it suffices to show that $\alpha \subseteq \gamma_x$. Let a be an arbitrary element of α. Since $i_y \varepsilon \beta$ we have $[(a \wedge i_x) \square i_y] \varepsilon \gamma$, and hence $[(a \wedge i_x) \square i_y]_x \varepsilon \gamma_x$. By P5 we then have $(a \wedge i_x) \varepsilon \gamma_x$; and hence, since $(a \wedge i_x) \leqq a$, $a \varepsilon \gamma_x$, as was to be shown.

The proof that $\gamma_y = \beta$ is similar.

LEMMA 22. *If* $\Lambda \subset A \subseteq I_x$ *and* $\Lambda \subset B \subseteq I_y$, *then* $(A \square B)_x = A$ *and* $(A \square B)_y = B$.

Proof. The inclusion $(A \square B)_x \subseteq A$ follows immediately from Lemma 21. To see that $A \subseteq (A \square B)_x$, suppose that $\alpha \varepsilon A$; we wish to show that $\alpha \varepsilon (A \square B)_x$. Since, by hypothesis, B is not empty, let $\beta \varepsilon B$. Since $\alpha \square \beta$ is not empty, let $\gamma \varepsilon (\alpha \square \beta)$; then $\gamma_x \varepsilon (\alpha \square \beta)_x$. Since $\alpha \varepsilon A$, and $A \subseteq I_x$, we see that $\alpha \varepsilon I_x$. Similarly, $\beta \varepsilon I_y$. Thus α, β, and γ satisfy the hypothesis of Lemma 21, so we conclude that $\alpha = \gamma_x$. Thus $\alpha \varepsilon (\alpha \square \beta)_x$. Since $(\alpha \square \beta)_x \subseteq (A \square B)_x$, we conclude finally that $\alpha \varepsilon (A \square B)_x$, as was to be shown.

The proof that $(A \square B)_y = B$ is similar.

LEMMA 23. *If* α *is a prime ideal, and if* $w \varepsilon \alpha$ *and* $v \varepsilon \alpha$, *then* $(u_x \square v_y) \varepsilon \alpha$.

Proof. Since $u \varepsilon \alpha$ and $v \varepsilon \alpha$, we have $(u \wedge v) \varepsilon \alpha$. By C10 we have $(u \wedge v) \leqq [(u \wedge v)_x \square (u \wedge v)_y]$, and by C4 and C9 we see that $[(u \wedge v)_x \square (u \wedge v)_y] \leqq (u_x \square v_y)$; hence $(u \wedge v) \leqq (u_x \square v_y)$, so that $(u_x \square v_y) \varepsilon \alpha$, as was to be shown.

LEMMA 24. *If* α *is any prime ideal, then* $\alpha \varepsilon (\alpha_x \square \alpha_y)$.

Proof. To prove the lemma it suffices to show that $(\alpha_x \square \alpha_y) \subseteq \alpha$, and hence it suffices to show that $(\alpha_x \boxtimes \alpha_y) \subseteq \alpha$. Let, then, w be an arbitrary element of $\alpha_x \boxtimes \alpha_y$. By Definition 5, there is an element u in α_x, and an element v in α_y such that $w = (u \wedge i_x) \square (v \wedge i_y)$. By Definition 2, there are then elements u_1 and v_1 in α such that $(u_1)_x \leqq u$ and $(v_1)_y \leqq v$. By Lemma 23 we see that $[(u_1)_x \square (v_1)_y] \varepsilon \alpha$. Since, by C2, $(u_1)_x \leqq i_x$ and $(v_1)_y \leqq i_y$, we have $(u_1)_x \leqq u \wedge i_x$ and $(v_1)_y \leqq v \wedge i_y$; hence, by C9, $[(u_1)_x \square (v_1)_y] \leqq w$, so that $w \varepsilon \alpha$, as was to be shown.

LEMM 25. *If* A, B, *and* C *are non-empty classes of prime ideals such that* $C_x = A$ *and* $C_y = B$, *then* $C \subseteq (A \square B)$.

Proof. Let α be any member of C. Then $\alpha_x \,\varepsilon\, A$ and $\alpha_y \,\varepsilon\, B$, and hence $(\alpha_x \,\Box\, \alpha_y) \subseteq (A \,\Box\, B)$. But, by Lemma 24, $\alpha \,\varepsilon\, (\alpha_x \,\Box\, \alpha_y)$. Hence $\alpha \,\varepsilon\, (A \,\Box\, B)$, as was to be shown.

THEOREM I. *The class of all classes of prime ideals of a projective algebra is a complete atomic projective algebra under the operations of union, intersection, complementation (with respect to the class, I, of all prime ideals) and the operations A_x, A_y, and $A \,\Box\, B$ introduced in Definitions* 3 *and* 6.

Proof. It is clear that the algebra is a complete atomic Boolean algebra. From Lemmas 5, 13, 6, 11, 22 and 25, 20, and 16 (respectively) we see that Everett and Ulam's postulates P1, $\cdots$, P7 are satisfied.

LEMMA 26. *If* $b \leqq a_x$ *and* $(b \,\Box\, i_y) \wedge a = 0$, *then* $b = 0$.

Proof. Since $(b \,\Box\, i_y) \wedge a = 0$, we have $a \leqq (b \,\Box\, i_y)'$; and hence, by C21, $a \leqq (b' \wedge i_x) \,\Box\, i_y$. Hence, by C1, $a_x \leqq [(b' \wedge i_x) \,\Box\, i_y]_x$. We then conclude by P5 that $a_x \leqq (b' \wedge i_x)$. Thus $b \leqq a_x \leqq (b' \wedge i_x) \leqq b'$; and hence $b = 0$, as was to be shown.

Definition 7. If a is any element of a projective algebra, then by $F(a)$ we shall mean the class of all prime ideals α such that $a \,\varepsilon\, \alpha$.

Remark. Thus, to say that $\alpha \,\varepsilon\, F(a)$ is equivalent to saying that $a \,\varepsilon\, \alpha$.

LEMMA 27. *If a is any element of a projective algebra, then* $F(a_x) = [F(a)]_x$.

Proof. Suppose first that $\alpha \,\varepsilon\, [F(a)]_x$. Then there exists a prime ideal β such that $\beta \,\varepsilon\, F(a)$ and $\alpha = \beta_x$. Then $a \,\varepsilon\, \beta$, and hence $a_x \,\varepsilon\, \beta_x$, or $a_x \,\varepsilon\, \alpha$. But $a_x \,\varepsilon\, \alpha$ means that $\alpha \,\varepsilon\, F(a_x)$. Hence $[F(a)]_x \subseteq F(a_x)$.

Now suppose, on the other hand, that $\alpha \,\varepsilon\, F(a_x)$. Let β be the set of all elements of the form $(u \,\Box\, i_y) \wedge a$, where $u \,\varepsilon\, \alpha$ and $u \leqq i_x$. β is not empty, since $[(a_x \,\Box\, i_y) \wedge a] \,\varepsilon\, \beta$.

Moreover, we notice that $0 \not\varepsilon\, \beta$. For otherwise there would be an element u of α such that $(u \,\Box\, i_y) \wedge a = 0$. Then, by C9, $[(u \wedge a_x) \,\Box\, i_y] \wedge a = 0$, and hence, by Lemma 26, $u \wedge a_x = 0$. Since $u \leqq i_x$, we have, by C8, $u_x = u$, and hence $u \,\varepsilon\, \alpha_x$. Since $\alpha \,\varepsilon\, F(a_x)$, we have $a_x \,\varepsilon\, \alpha$, and hence, making use of P4, $a_x \,\varepsilon\, \alpha_x$. Thus $(u \wedge a_x) \,\varepsilon\, \alpha_x$, or $0 \,\varepsilon\, \alpha_x$, which is absurd in view of the fact that α_x is a prime ideal.

We notice next that if $u_1 \,\varepsilon\, \beta$ and $u_2 \,\varepsilon\, \beta$ then $(u_1 \wedge u_2) \,\varepsilon\, \beta$. For suppose that $u_1 = [(v_1 \,\Box\, i_y) \wedge a]$ and $u_2 = [(v_2 \,\Box\, i_y) \wedge a]$, where $v_1 \,\varepsilon\, \alpha$, $v_2 \,\varepsilon\, \alpha$,

$v_1 \leqq i_x$ and $v_2 \leqq i_x$. Then, making use of C11, we have $u_1 \wedge u_2 = [(v_1 \wedge v_2) \square i_y] \wedge a$, and hence $(u_1 \wedge u_2) \varepsilon \beta$, as was to be shown.

We now define γ to be the set of all elements u such that, for some element v of β, $v \leqq u$. From what has been proved about β, it is easily seen that γ is an ideal. Let δ be any prime ideal such that $\gamma \subseteq \delta$.

Since $a_x \varepsilon \alpha$, we see that $i_x \varepsilon \alpha$, and hence $[(i_x \square i_y) \wedge a] \varepsilon \beta$. Hence, by C10, we conclude that a is in β, and hence in δ. Hence $\delta \varepsilon F(a)$. To complete our proof, it will therefore suffice to show that $\alpha = \delta_x$. Since δ_x and α are both prime ideals, however, it is sufficient to show that $\alpha \subseteq \delta_x$.

Suppose then that $u \varepsilon \alpha$. Since $i_x \varepsilon \alpha$, we have $(u \wedge i_x) \varepsilon \alpha$. Hence $[(u \wedge i_x) \square i_y] \wedge a$ is in β, and hence in δ. Hence $(u \wedge i_x) \square i_y$ is in δ, and hence $[(u \wedge i_x) \square i_y]_x$ is in δ_x. Hence, by P5, $(u \wedge i_x)$ is in δ_x, and hence u is in δ_x, as was to be shown.

LEMMA 28. *If a is any element of a projective algebra, then* $F(a_y) = [F(a)]_y$.

Proof. Similar to the proof of Lemma 27.

LEMMA 29. *If a_1 and a_2 are any elements of a projective algebra (with $a_1 \leqq i_x$ and $a_2 \leqq i_y$), then $F(a_1 \square a_2) = F(a_1) \square F(a_2)$.*

Proof. If $a_1 = 0$, then, using C7, $F(a_1 \square a_2) = F(0) = \Lambda = \Lambda \square F(a_2) = F(a_1) \square F(a_2)$. Similarly if $a_2 = 0$. Hence we shall take $a_1 \neq 0 \neq a_2$.

Now suppose first that $\alpha \varepsilon F(a_1) \square F(a_2)$. Then there is an $\alpha_1 \varepsilon F(a_1)$ and an $\alpha_2 \varepsilon F(a_2)$ such that $\alpha \varepsilon (\alpha_1 \square \alpha_2)$; and hence such that $(\alpha_1 \boxtimes \alpha_2) \subseteq \alpha$. Since $a_1 \varepsilon \alpha_1$ and $a_2 \varepsilon \alpha_2$, we therefore conclude that $(a_1 \square a_2) \varepsilon \alpha$, and hence that $\alpha \varepsilon F(a_1 \square a_2)$, as was to be shown.

Now suppose, on the other hand, that $\alpha \varepsilon F(a_1 \square a_2)$. Then $(a_1 \square a_2) \varepsilon \alpha$, and hence $(a_1 \square a_2)_x \varepsilon \alpha_x$, or $a_1 \varepsilon \alpha_x$. Similarly $a_2 \varepsilon \alpha_y$. Thus $\alpha_x \varepsilon F(a_1)$ and $\alpha_y \varepsilon F(a_2)$; and hence, by Definition 6, $(\alpha_x \square \alpha_y) \subseteq [F(a_1) \square F(a_2)]$. By Lemma 24, on the other hand, $\alpha \varepsilon (\alpha_x \square \alpha_y)$. Hence we conclude that $\alpha \varepsilon [F(a_1) \square F(a_2)]$, as was to be shown.

THEOREM II. *Every projective algebra is isomorphic to a subalgebra of a complete atomic projective algebra.*

Proof. Let K be any projective algebra, and let Γ be the corresponding algebra of prime ideals over K. By Theorem I, Γ is a complete atomic projective algebra. Let Δ be the class of elements A of Γ such that, for

some a in K, $A = F(a)$. It follows from the results of M. H. Stone [8] that whenever $F(a) = F(b)$ then $a = b$, and that

$$F(a \wedge b) = F(a) \cap F(b)$$
$$F(a \vee b) = F(a) \cup F(b)$$
$$F(a') = I - F(a).$$

Moreover, from Lemmas 27, 28, and 29 we have

$$F(a_x) = [F(a)]_x$$
$$F(a_y) = [F(a)]_y$$
$$F(a \square b) = F(a) \square F(b).$$

Thus Δ is a subalgebra of Γ, and is isomorphic to K.

From Theorem II, together with the results of Everett and Ulam, we obtain the following.

THEOREM III. *Every abstract projective algebra is isomorphic to a projective algebra of subsets of a direct product.*

OKLAHOMA A. & M. COLLEGE.

THE ARITHMETIC OF BELL AND STIRLING NUMBERS.*

By H. W. BECKER and JOHN RIORDAN.

1. Introduction. The Stirling numbers of the title are those introduced by E. T. Bell [3] and are derived from ordinary numbers of the second (or first) kind by repeated matrix multiplication: thus

$$(1.1) \qquad S(c, r, s) = \Sigma_i\, S(c, i, s-t) S(i, r, t)$$

where the summation is over all values for which the numbers exist. The Bell numbers introduced in Bell [2] are Stirling numbers summed on c, i. e.:

$$(1.2) \qquad B(r, s) = \Sigma_c\, S(c, r, s) = \Sigma_i\, S(i, r, t) B(i, s-t)$$

the last as a consequence of (1.1).

The notation adopted here, which is an innovation, is intended to suggest the exhibition of the Stirling numbers in three dimensional lattices, c for column, r for row and s for stack or index.

The interesting thing, aesthetically, in the arithmetic of these numbers is their periodicity modulo p, a prime. For the Bell numbers of index 1, fairly complete results are known and are reviewed in **2**. In **3** we give a generalization of these to numbers of index s, proving that the period is

$$p^{p^m} - 1, \qquad p^{m-1} \leq s < p^m.$$

Similar results for the simpler Stirling numbers (index 1) which can hardly be new are given in **4**, while **5** treats the general case.

Our treatment, following Bell, rests essentially on the Lagrange identical congruence

$$(1.3) \qquad (x)_p \equiv x^p - x,\ (p),$$

$(x)_p$ being the Jordan, [6] and [7], factorial notation;

$$(x)_p = x(x-1) \cdots (x-p+1) = \Sigma_i\, S(i, p, -1) x^i.$$

Of its many extensions, Bell [2], we need only the following

$$(1.4) \qquad (x^p - x)^m x^r \equiv (x^p - x)^m\, \Sigma_i\, S(i, r, 1)(x)_i,$$
$$\equiv \Sigma_i\, S(i, r, 1)(x)_{mp+i}.$$

* Received June 25, 1947.

Here and from now on all congruences are taken modulus p, a prime, without explicit indication.

For use of these in the sequel, it is worth noting that the instances $t = -1$ of (1.1) and (1.2) may be written more concisely as

$$S(c, r, s) = [S(c, \ , s+1)]_r, \tag{1.1a}$$

$$B(r, s) = [B(\ , s+1)]_r, \tag{1.2a}$$

In these the terms on the right are to be interpreted symbolically using the conventions of the Blissard or umbral calculus: $x^i \equiv x_i$, $x^0 \equiv x_0$, with blank spaces indicating the variable so treated; thus

$$\begin{aligned}[S(c, \ , s+1)]_r &= \Sigma_i S(i, r, -1)\,[S(c, \ , s+1)]^i, \\ &= \Sigma_i S(i, r, -1)\, S(c, i, s+1).\end{aligned}$$

For concreteness, a table of the Stirling numbers of general index for the first few rows and columns appears below.

TABLE I.

Stirling Numbers $S(c, c+r, s)$

c/r	0	1	2	3	4
1	1	s	$s(3s-1)/2$	$s(2s-1)(3s-1)/2$	$s(45s^3-65s^2+30s-4)/6$
2	1	$3s$	$s(9s-2)$	$5s(3s-1)(4s-1)/2$	$3s(225s^3-235s^2+80s-8)/6$
3	1	$6s$	$5s(6s-1)$	$15s(4s-1)(5s-1)/2$	$21s(225s^3-185s^2+50s-4)/6$
4	1	$10s$	$5s(15s-2)$	$35s(5s-1)(6s-1)/2$	$14s(1575s^3-1070s^2+240s-16)/6$
5	1	$15s$	$35s(9s-1)/2$	$70s(6s-1)(7s-1)/2$	$126s(630s^3-365s^2+70s-4)/6$

$$\begin{aligned}&S(c, c+1, s) = {}_{c+1}C_2 s; \qquad S(c, c+2, s) = {}_{c+2}C_3 s[3(c+1)s-2]/4 \\ &S(c, c+3, s) = {}_{c+}C_4 s[(c+1)s-1][(c+2)s-1]/2 \\ &S(c, c+4, s) = {}_{c+4}C_5 s[15(c+3)_3 s^3 - 20(3c^2+12c+11)s^2 \\ &\qquad\qquad\qquad\qquad + 80(c+2)s - 32]/48.\end{aligned}$$

For any s, $S(c, r, s) = 0$, $c > r$, $c < 0$ or $r < 0$; $S(c, c, s) = 1$, $S(0, r, s) = 0$, $r \neq 0$, and $S(c, r, 0) = \delta(c, r)$, the Kronecker delta. Similar

results for Bell numbers, which are used later, are obtained by summing on diagonals; thus $B(1,s)=1$, $B(2,s)=1+s$, and so on.

2. Exponential integers. A brief résumé of the arithmetic of the Bell numbers of index 1, due mainly to Bell [2], Touchard [8] and Hall [4], but rediscovered by Williams [9], is given here for the light it casts on the general case.

From

$$[B(\ ,1)]_r = B(r,0) = 1$$

which is an instance of (1.2a), and (1.4) with $m=1$ and $x^r \equiv B(r,1)$, the Touchard congruence

$$B(p+r,1) \equiv B(r+1,1) + B(r,1) \tag{2.1}$$

follows at once. Regarded as a difference congruence, this has the characteristic equation [1]

$$E_1{}^p \equiv E_1 + 1, \tag{2.2}$$

E_1 of course being a shift operator such that $E_1B(r,1) = B(r+1,1)$.

Congruence (2.2) implicitly contains the whole of the arithmetic of these numbers since

$$E_1{}^{p^2} \equiv (E_1+1)^p \equiv E_1{}^p + 1 \equiv E_1 + 2,$$

$$E_1{}^{p^i} \equiv E_1 + i, \tag{2.3}$$

$$E_1{}^{ap^i} \equiv (E_1+i)^a, \qquad a < p, \tag{2.4}$$

$$E_1{}^{[p]}) \equiv \Pi(E_1+i)^{a_i}, \qquad [p] = \Sigma_i a_i p^i. \tag{2.5}$$

Writing $B(r,1) \equiv B_r$ for brevity, the last corresponds to

$$B_{r+[p]} \equiv \Pi(B+i)^{a_i} B^r, \tag{2.6}$$

where the right hand side is symbolic.

If $a_i = 1$ for $i = 0$ to $p-1$,

$$[p] = 1 + p + p^2 + \cdots p^{p-1} = (p^p - 1)(p-1)^{-1} = P,$$

[1] Thanks are due Marshall Hall for a long communication to one of us showing the elegance of the characteristic equation in deriving the period of the exponential numbers given in his unpublished paper [4].

and

$$(27)\qquad \begin{aligned} B_{r+P} &\equiv B^{r+1}(B+1)(B+2)\cdots(B+p-1),\\ &\equiv B^r(B^p-B),\\ &\equiv B_r, \end{aligned}$$

which is Hall's result (rediscovered by G. T. Williams [9]) that P, or some divisor of P of course, is the period of the exponential numbers. The result used in the second line is a consequence of Lagrange's congruence.

3. Bell numbers. By (1.2a) and the instance $m=1$ of (1.4) with x replaced by the umbra $B(\ ,s+1)$:

$$(3.1)\qquad B(p+r,s+1)-B(r+1,s+1)\equiv\Sigma_i\, S(i,r,1)B(p+i,s).$$

The characteristic equation corresponding to this is

$$(3.2)\qquad E_{s+1}{}^p = E_{s+1} = E_s{}^p$$

where

$$E_jB(r,s)=\Sigma_iB(i+1,j)S(i,r,s-j);$$

that is to say, the shift operator E_i operates on the Bell number of corresponding index in the second form of equation (1.2).

Equation (3.2) with the boundary relation: $E_0=1$ leads by iteration to

$$(3.3)\qquad E_s{}^p\equiv E_s+E_{s-1}+\cdots+E_1+1,$$

so that

$$\begin{aligned} E_s{}^{p^2} &\equiv E_s{}^p+E_{(s-1)}{}^p+\cdots+E_1{}^p+1,\\ &\equiv E_s+2E_{s-1}+\cdots+(j+1)E_{s-j}+\cdots+sE_1+s+1, \end{aligned}$$

and finally, provable by induction:

$$(3.4)\qquad E_s{}^{p^i}\equiv E_s+iE_{s-1}+\cdots+\binom{i+j-1}{j}E_{s-j}+\cdots+\binom{i+s-1}{s}.$$

The formula for the congruence of a general Bell number, index s, with r arbitrary in the scale of p is similar to (2.5)—with E_1+i replaced by the right of (3.4)—and need not be written out in full. But we should note, for verifications, the congruence corresponding to (3.4) namely:

$$(3.5)\qquad B(p^i+r,s)\equiv\sum_j\sum_k\binom{i+j-1}{j}B(k+1,s-j)S(k,r,j).$$

For $r=0$, $S(k,0,j)=1$, $k=0$, and zero otherwise, so that

$$(3.6)\quad B(p^i, s) \equiv \sum_j \binom{i+j-1}{j} B(1, s-j) = \sum_{j=0}^{s} \binom{i+j-1}{j} = \binom{i+s}{s},$$

in agreement with Bell [2], 4. 5.

For $r = 1$, $S(0, 1, j) = 0$, $S(1, 1, j) = 1$, and

$$(3.7)\qquad B(p^i + 1, s) \equiv \Sigma_j \binom{i+j-1}{j} B(2, s-j),$$

$$= \Sigma_j \binom{i+j-1}{j} (s-j+1) = \binom{i+s+1}{s}.$$

in agreement with Bell [2] eq. 4. 6. The next three results of the same kind (not given explicitly by Bell) are as follows:

$$(3.8)\qquad B(p^i + 2, s) \equiv B(2, s) \binom{i+s+1}{s} + \binom{i+s+1}{s-1},$$

$$(3.9)\qquad B(p^i + 3, s) \equiv B(3, s) \binom{i+s+1}{s} + (3s+2) \binom{i+s+1}{s-1},$$

$$(3.10)\quad B(p^i + 4, s) \equiv B(4, s) \binom{i+s+1}{s}$$

$$+ (9s^2 + 10s + 3) \binom{i+s+1}{s-1} + \binom{i+s+1}{s-2},$$

where

$$B(2, s) = s + 1,$$

$$B(3, s) = s + 1 + 3 \binom{s+1}{2},$$

$$B(4, s) = s + 1 + 13 \binom{s+1}{2} + 18 \binom{s+1}{3}.$$

A more interesting use of (3. 4) is in determining the period of the numbers. Note first that the coefficients on its right are those in

$$(1-x)^{-i} = \Sigma_j \binom{i+j-1}{j} x^j.$$

Hence for $i = p$, it may be reduced by the congruence (Hardy and Wright [5] Theorem 76)

$$(1-x)^{-p} \equiv 1 + x^p + x^{2p} + \cdots.$$

For $s < p$ and $q = p^p$,

$$E_s{}^q \doteq E_s,$$

and

$$(3.11)\qquad B(q + r - 1, s) \equiv B(r, s).$$

The period of $B(r, s)$ is then $p^p - 1$ for all s less than p. $B(r, 1)$ of course has the smaller period, 2, $(p^p - 1)(p - 1)^{-1}$. For specific values of p, smaller periods may appear. We have no results on these but have verified that $B(r, 2)$ has its full period 26 for $p = 3$.

For $p \leq s < p^2$, and

$$E_{s^q} \equiv E_s + E_{s-p} + E_{s-2p} + \cdots,$$

$$E_{s^{q^p}} \equiv E_s,$$

by a repetition of the preceding argument, so that the period for $p \leq s < p^2$ is $p^{p^2} - 1$.

The general result for the period is

$$P = p^{p^m} - 1, \qquad p^{m-1} \leq s < p^m. \tag{3.12}$$

This may be proved as follows. First the persistence of the period over the interval is evident from the results given above. In (3.4) put $i = s = p^m$; then since

$$\binom{2p^m - j - 1}{p^m - 1} \equiv 1, j = 0, p^m,$$
$$\equiv 0 \text{ otherwise,}$$

all terms of (3.4) vanish, save the first and last, and

$$E^Q{}_{p^m} \equiv E_{p^m} + 1, \quad Q = p^{p^m}, \tag{3.13}$$

or

$$B(Q + r, p^m) \equiv B(r + 1, p^m) + B(r, p^m). \tag{3.14}$$

This reduces to Touchard's congruence (2.1) for $m = 0$. Also

$$E^{Q^p}{}_{p^m} \equiv E_{p^m}$$

or

$$B(Q^p + r - 1, p^m) \equiv B(r, p^m). \tag{3.15}$$

Hence the period for $s = p^m$ is $Q^p - 1$ or $p^{p^{m+1}} - 1$, as in (3.12).

4. Stirling numbers of the second kind. Though strikingly similar to that of Bell numbers, the arithmetical structure of the generalized Stirling numbers is sufficiently different to warrant a preliminary study of a simple case, the Stirling numbers of the second kind, for orientation.

Using (1.4) with $m = 1$ and x^r replaced by $S(c, r, 1)$, or for brevity $S(c, r)$, and the instance $s = 0$ of (1.1a), a congruence which sums to the Touchard congruence, (2.1) above, is obtained at once:

(4.1) $$S(c, p+r) \equiv S(c, r+1) + \Sigma_i S(i, r)\delta(c, p+i),$$
or
$$S(c, p+r) \equiv S(c, r+1) + S(c-p, r).$$

Note that $S(c, r) = 0$, $c < 0$.

Then

(4.1a) $$S(c, p+r) \equiv S(c, r+1), \quad c < p,$$

and in this range of c the numbers have period $p-1$. For other values of c it is convenient to replace r by $c+r$ since $S(c, r) = 0$, $r < c$; this squares off the triangular matrix of values. From (4.1) it is apparent that the residues may be treated in rectangular blocks of p by $p-1$ and each block is the sum of the block above and the block on its left; hence

(4.2) $$S(c+ip, c+ip+r+j(p-1)) \equiv \binom{i+j}{i} S(c, c+r).$$

Here i and j are positive integers or zero and the ranges of c and r are 1 to p and 0 to $p-2$, respectively. The arithmetical structure then follows from that of the binomial coefficients.

The characteristic equation corresponding to (4.1) is

(4.3) $$E_1^p \equiv E_1 + E_0^p$$

with both operators working only on the row index as with Bell numbers:

$$E_0 S(c, r) = \Sigma S(i, r)\delta(c, i+1) = S(c-1, r).$$

By iteration of (4.3),

(4.4) $$E_1^{p^i} \equiv E_1 + E_0^p + E_0^{p^2} + \cdots + E_0^{p^i},$$
so that
$$S(c, r+p^i) \equiv S(c, r+1) + S(c-p, r) + \cdots + S(c-p^i, r).$$

Note that the left has $i+1$ terms in concordance with the Bell number congruence $E_1^{p^i} \equiv E_1 + i$. For $c < p$, all but the first term on the left are vacuous; for $p \leq c < p^2$ all but the first two are vacuous and so on. It follows at once from this that the period of the numbers is

(4.5) $$P = p^j(p-1), \quad p^j \leq c < p^{j+1},$$

for in this range of c both $S(c, r+p^j)$ and $S(c, r+p^{j+1})$ have the same congruence. This may be improved slightly since, from (4.2), the period for $c = p^j$ is $p^{j-1}(p-1)$.

5. Generalized Stirling numbers. For the generalized Stirling number $S(c,r,s)$, the fundamental characteristic equation is

$$E_s{}^p \equiv E_s + E_{(s-1)}{}^p. \tag{5.1}$$

By iteration, and equation (4.3), this leads at once to

$$E_s{}^p \equiv E_s + E_{s-1} + \cdots + E_1 + E_0{}^p, \tag{5.2}$$

and to the more general result

$$\begin{aligned} E_s{}^{p^i} \equiv E_s + iE_{s-1} + \cdots + \binom{i+j-1}{j} E_{s-j} + \cdots + \binom{i+s-2}{s-1} E_1 \\ + \binom{i+s-2}{s-2} E_0{}^p + \cdots + \binom{i+s-1-j}{s-1} E_0{}^{p^j} + \cdots + E_0{}^{p^i} \end{aligned} \tag{5.3}$$

This congruence plays the same role in the arithmetic of Stirling numbers that (3.4) does for Bell numbers. In non-symbolic form it reads:

$$\begin{aligned} S(c,p^i+r,s) \equiv \sum_{j=0}^{s-1} \sum_k \binom{i+j-1}{j} S(k,r,j) S(c,k+1,s-j) \\ + \sum_{j=1}^{i} \binom{i+s-1-j}{s-1} S(c-p^j,r,s). \end{aligned} \tag{5.4}$$

Hence it may be used for specific results and for periodicity.

For the former note first the instance $r=0$, for which

$$S(1,p^i,s) \equiv \binom{i+s-1}{s-1}, \tag{5.5}$$

$$S(c,p^i,s) \equiv \sum_{j=1}^{i} \binom{i+s-1-j}{s-1} \delta(c,p^j), \quad c>1. \tag{5.6}$$

As instances of the last, verified by Bell ([3], equations 4.2, 4.14, 4.15, respectively) note that

$$S(c,p^i,s) \equiv 0, \quad 1<c<p,$$

$$S(c,p,s) \equiv \delta(c,p),$$

$$S(c,p^2,s) \equiv s\delta(c,p) + \delta(c,p^2).$$

It is convenient to exhibit results for two conditions: c and p greater than $r+1$ and less than $r+1$. For the former

$$S(c,p^i+r,s) \equiv \sum_{j=1}^{i} \binom{i+s-1-j}{s-1} S(c-p^j,r,s). \tag{5.7}$$

For the latter the results corresponding to those given for Bell numbers may be conveniently exhibited in the table below, where for brevity B^*_r is written for $B(p^i+r, s-1)$ and S_{ij} for $S(i,j,s)$

r/c	1	2	3	4
0	B^*_0			
1	B^*_1	B^*_0		
2	B^*_2	$2B^*_1+S_{12}B^*_0$	B^*_0	
3	B^*_3	$3B^*_2+S_{23}B^*_1+S_{13}B^*_0$	$3B^*_1+S_{23}B^*_0$	B^*_0

Turning now to the period, results similar to those for Bell numbers (and in form similar to those of Stirling numbers of the second kind) may be obtained by a similar method. We first state the general result: writing $q_m = p^{p^m}$, then the period for the general Stirling number $S(c,r,s)$ is

$$(5.8) \qquad P_{c,s} = q_m{}^i(q_m-1), \qquad p^{m-1} < s \leq p^m, \quad q_m{}^{i-1}p^s < c < q_m{}^i p^s.$$

For orientation, the case $s=2$ deserves notice as the simplest available. Here, by equation (5.3),

$$(5.9) \qquad E_2{}^{q_1} \equiv E_2 + (p-1)E_0{}^{p^2} + \cdots + (p+1-j)E_0{}^{pj} + \cdots + E_0{}^{q_1},$$

and

$$E_2{}^{q_1} \equiv E_2, \quad c < p^2,$$

since all terms after the first are vacuous in the given range of c. For $p^2 < c < q_1 p^2$,

$$(5.10) \qquad E_2{}^{q_1^2} \equiv E_2{}^{q_1},$$

for the same reason. Continuation of the process leads to

$$(5.11) \qquad E_2{}^{q_1{}^i} \equiv E_2{}^{q_1{}^{i-1}}, \qquad q_1{}^{i-1}p^2 < c < q_1{}^i p^2.$$

Now put $i = s = p^m$ in equation (5.3); then

$$(5.12) \qquad E_{p^m}{}^{q_m} \equiv E_{p^m} + E_0{}^{q_m},$$

by the binomial congruences used for the Bell numbers. By the process used above, it appears at once that

$$(5.13) \qquad E_{p^m}{}^{q_m{}^i} \equiv E_{p^m}{}^{q_m{}^{i-1}}, \qquad q_m{}^{i-1} < c < q_m{}^i,$$

in accordance with (5.8).

For $i = p^m$, $s = p^m + 1$,

$$E_s{}^{q_m} \equiv E_s + E_1{}^{q_m}, \tag{5.14}$$

and for $i = p^m$, $s = p^m + k$

$$E_s{}^{q_m} \equiv E_s + E_k{}^{q_m}, \tag{5.15}$$

and (5.8) follows from a detailed application of the process illustrated for $s = 2$.

DESCANSO, CALIFORNIA.
BELL TELEPHONE LABORATORIES,
NEW YORK.

BIBLIOGRAPHY

1. E. T. Bell, "Exponential polynomials," *Annals of Mathematics*, vol. 35 (1934), pp. 258-277.
2. ———, "The iterated exponential integers," *Annals of Mathematics*, vol. 39 (1938), pp. 539-557.
3. ———, "Generalized Stirling transforms of sequences," *American Journal of Mathematics*, vol. 61 (1939), pp. 89-101.
4. Marshall Hall, "Arithmetic properties of a partition function," (Abstract 200) *Bulletin of the American Mathematical Society*, vol. 40 (1934).
5. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford, 1938.
6. C. Jordan, *Statistique Mathématique*, Paris, 1927.
7. ———, "On Stirling's numbers," *The Tohoku Mathematical Journal*, vol. 37 (1933), pp. 254-278.
8. J. Touchard, "Propriétés arithmétiques de certains nombres recurrents," *Annales de la Société Scientifique de Bruxelles*, A53 (1933), pp. 21-31.
9. G. T. Williams, "Numbers generated by the function e^{e^x-1}," *American Mathematical Monthly*, vol. 52 (1945), pp. 323-327.

ON A THEOREM OF MILLOUX.*

By Philip Hartman.

1. In the system of linear, homogeneous differential equations,

$$z' = A(t)z, \qquad (' = d/dt), \tag{1}$$

let t denote a real variable, $A(t)$ an n by n matrix of functions (possibly complex-valued) and z a vector with n components. An asterisk will denote the complex-conjugate transpose for square matrices, vectors or scalars. The main theorem to be proved is

(i) *If $A(t)$, where $0 \leqq t < \infty$, is a continuous matrix function having the property that*

$$\lim_{t\to\infty} |z(t)| \text{ exists for every solution } z = z(t) \tag{2}$$

of (1), *then necessary and sufficient for the existence of at least one solution vector $z = z(t) \not\equiv 0$ satisfying*

$$\lim_{t\to\infty} |z(t)| = 0 \tag{3}$$

is

$$\Re \int_0^t \text{trace } A(s)\, ds \to -\infty \text{ as } t \to \infty. \tag{4}$$

It is possible to state criteria in terms of the coefficient matrix $A(t)$ which assure that the system (1) has the property (2). In this direction, a result of Wintner [2], pp. 557-559, on non-linear systems implies

(ii) *If $A(t)$, where $0 \leqq t < \infty$, is a continuous matrix function such that the Hermitian matrix*

$$A(t) + A^*(t) \text{ is non-positive definite for } 0 \leqq t < \infty, \tag{5}$$

then every solution vector $z = z(t)$ of (1) *satisfies*

$$|z(t)| \text{ is non-increasing;} \tag{6}$$

hence (2) *holds.*

While conditions (4) and (5) together imply the existence of at least

* Received January 10, 1948.

one non-trivial solution of (1) satisfying (3), they do not imply that *all* solutions of (1) satisfy (3), see **2** below. If, however, the "averaged condition" (4) is replaced by

$$\int^{t} \max_{|z|=1} z^{*}\{A(s)+A^{*}(s)\}z\,ds \to -\infty \text{ as } t\to\infty,$$

then every solution of (1) satisfies (3); cf. Wintner [2], p. 557.

2. The assertions (i) and (ii) imply the following theorem of Milloux [1], p. 49:

(iii) *If $f(s)$ is a real-valued, continuous (scalar) function which, for large s, is monotone and satisfies*

(7) $$f(s)\to\infty \text{ as } s\to\infty,$$

then the differential equation

(8) $$\ddot{x}+f(s)x=0 \qquad (\dot{} = d/ds)$$

possesses at least one non-trivial solution $x=x(s)\not\equiv 0$ satisfying

(9) $$x(s)\to 0 \text{ as } s\to\infty.$$

The question of the existence of such a solution was first raised by Biernacki, cf. Milloux [1], p. 49. The proofs of (i) and (ii) furnish a simple proof of (iii), in contrast to the very complicated proof of Milloux.

An example given by Milloux [1], p. 50, shows that not every solution of (8) must satisfy (9). Although, in his example, the coefficient function $f(s)$ is a monotone step-function, this function can be modified [1] so as to

[1] When $f(s)$ possesses isolated discontinuity points, by a solution $x=x(s)$ of (8) for $s_0<s<\infty$ is meant a function which possesses, on $s_0<s<\infty$, a continuous first derivative and, at every continuity point of $f(s)$, a second derivative, which is continuous and satisfies (8).

The required modification of Milloux's example $f(s)$ can be obtained as follows: The non-decreasing property of $f(s)$ implies that every solution $x=x(s)$ of (8) is bounded (as $s\to\infty$). It follows, therefore, that if $g(s)$ is a function (with isolated discontinuities) satisfying

$$\int^{\infty} |g(s)|\,ds<\infty,$$

then every solution $x=x(s)$ of (8) satisfies (9) if and only if every solution $y=y(s)$ of

(8 bis) $$\ddot{y}+(f(s)+g(s))y=0$$

tends to 0 as $s\to\infty$; Wintner, [3], (iv_0), p. 267. (This theorem is stated and proved

furnish a smooth $f(s)$. The modified example shows that (4) and (5) do not imply that every solution of (1) must satisfy (3); cf. the deduction of (iii) from (i) and (ii) below.

3. *Proof of* (ii) *and* (i). In order to prove (ii), suppose that (5) holds. If $z = z(t)$ denotes an arbitrary solution vector of (1), then the derivative of $|z|^2 = z^*z$ is

$$z^*z' + z^{*\prime}z = z^*(A + A^*)z.$$

Hence, by the assumption (5), the derivative of the non-negative function $|z(t)|^2$ is non-positive. This proves (6) and (ii).

In order to prove (i), suppose that (2) holds. Let $Z(t)$ denote a fundamental matrix of (1), that is, a non-singular matrix in which the j-th column is a solution vector $z = z_j(t)$ of (1) for $j = 1, \cdots, n$. If c is an arbitrary constant vector, then the vector

$$z = z(t) = Z(t)c \tag{10}$$

is a solution of (1); furthermore, (10) is a non-trivial solution whenever $c \neq 0$. The squared length of (10) is the positive definite Hermitian form

$$|z(t)|^2 = c^*Z^*(t)Z(t)c. \tag{11}$$

The elements of the matrix $Z^*(t)Z(t)$ are the scalar products $z_j^*(t)z_k(t)$. By applying (2) to the solutions $z_j(t) + z_k(t)$ and $z_j(t) + iz_k(t)$ of (1), it is seen that the scalar products $z_j^*(t)z_k(t)$ tend to limits as $t \to \infty$. Hence, the Hermitian matrix $Z^*(t)Z(t)$ tends to a limit matrix, say H, and the corresponding limit belonging to the function (10) is the non-negative Hermitian form

$$c^*Hc = \lim_{t\to\infty} |z(t)|^2. \tag{12}$$

The proof of (i) will be complete if it is shown that (4) is necessary and sufficient for the existence of a vector $c \neq 0$ for which (12) vanishes. Since H is a non-negative definite Hermitian matrix, the existence of such a vector c is equivalent to the vanishing of the determinant of H. Clearly,

---

loc. cit. for the case where $f(s)$, $g(s)$ are continuous, but it is clear that the proof also holds for the type of functions at hand.) Hence, (8 bis) possesses a solution $y = x(s)$ violating (9), since the same is true of (8). Now, if $f(s)$ is the function in the example of Milloux, an absolutely integrable function $g(s)$, with a sequence of isolated discontinuity points, can be constructed such that $f(s) + g(s)$ is monotone, tends to infinity and is continuous (or even differentiable).

$$\det H = \lim_{t\to\infty} \det\{Z^*(t)Z(t)\} = \lim_{t\to\infty} |\det Z(t)|^2.$$

Hence, the Jacobi identity for fundamental matrices,

$$\det Z(t) = \det Z(0) \exp \int_0^t \operatorname{trace} A(s)\,ds, \qquad (\det Z(0) \neq 0),$$

shows that (4) is necessary and sufficient for $\det H = 0$.

This completes the proof of (i).

4. *Proof of* (iii). Let $s = s_0$ be chosen so large that $f(s)$ is positive and monotone non-decreasing for $s > s_0$. In what follows, only s-values satisfying $s \geqq s_0$ occur.

Suppose, for a moment, that $f(s)$ possesses a continuous derivative, $\dot{f}(s) \geqq 0$. If the new independent variable

$$t = t(s) = \int_{s_0}^{s} f^{\frac{1}{2}}(s)\,ds \tag{13}$$

is introduced, the differential equation (8) becomes

$$x'' + (\dot{f}/f^{3/2})x' + x = 0,$$

where the prime denotes differentiation with respect to t and the argument of the function occurring in the coefficient of x' is $s = s(t)$, the inverse of (13). The last differential equation is equivalent to a system (1), where n is 2, the vector z is (x, x') and the coefficient matrix is

$$A(t) = \begin{pmatrix} 0 & 1 \\ -1 & -\dot{f}/f^{3/2} \end{pmatrix}; \text{ hence, } A(t) + A^*(t) = \begin{pmatrix} 0 & 0 \\ 0 & -2\dot{f}/f^{3/2} \end{pmatrix}.$$

What remain to be verified are conditions (4) and (5) of (i) and (ii), respectively. Clearly, (5) holds, since $\dot{f} \geqq 0$. Also the integral in (4) is

$$-\int^{t} (\dot{f}/f^{3/2})\,dt = -\int^{s} (\dot{f}/f)\,ds = -\log f(s) + \text{const.},$$

by (13). Hence, (7)' implies (4). This proves (iii) if $f(s)$ possesses a continuous derivative.

If $f(s)$ does not have a continuous derivative, this proof can be modified along the following lines: Let $x = x(s)$ be any real solution of (8) and let $x_0(s)$ denote the "conjugate energy,"

$$x_0(s) = x^2(s) + \{\dot{x}^2(s)/f(s)\}. \tag{14}$$

Then, in virtue of (8) and the monotony of $f(s)$,

$$dx_0(s) \doteq \dot{x}^2(s)\,d(1/f(s)) \leq 0,$$

where the differentials are to be interpreted in the sense of Stieltjes integration; cf. Wintner [3], p. 257. Since (14) is non-negative, it follows that

$$\lim_{s\to\infty} x_0(s)$$

exists and is non-negative for every real solution $x = x(s)$ of (8). Clearly, the proof of (i) can now be applied to the present situation, where the vector z becomes $(x, \dot{x}/f^{\frac{1}{2}})$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] H. Milloux, "Sur l'équation différentielle $x'' + A(t)x = 0$," *Prace Matematyczno-Fizyczne*, vol. 41 (1934), pp. 39-53.

[2] A. Wintner, "Asymptotic integration constants," *American Journal of Mathematics*, vol. 68 (1946), pp. 553-559.

[3] ———, "Asymptotic integrations of the adiabatic oscillator, *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

THE SET ON WHICH AN ENTIRE FUNCTION IS SMALL.*

By R. P. Boas, Jr., R. C. Buck, and P. Erdös.

Let $f(z)$ be an entire function and $M(r)$ the maximum of $|f(z)|$ on $|z| = r$. We give some results on the density of the set of points at which $|f(z)|$ is small in comparison with $M(r)$; although simple, these results seem not to have been noticed before.

If E is a measurable set in the z-plane, we denote by $D_R(E)$ the ratio $m(z \varepsilon E, |z| \leq R)/\pi R^2$ and by $\bar{D}(E)$ and $\underline{D}(E)$ the upper and lower densities of E, that is the superior and inferior limits of $D_R(E)$ as $R \to \infty$. For a fixed function $f(z)$, let E_λ be the set of points z for which $\log |f(z)| \leq (1 - \lambda) \log M(|z|)$. Our results may be stated as follows.

THEOREM 1. *For any $\lambda > 1$, there is a number K, the same for all functions $f(z)$, such that $\bar{D}(E_\lambda) \leq K$. Moreover, $0 < K \leq \lambda^{-1}$.*

In particular, for $\lambda = 2$, the upper density of the set where $|f(z)| \leq 1/M(|z|)$ is at most 1/2. Much stronger results are known for entire functions of small finite order. The interest of Theorem 1 is that it holds for all entire functions and that, contrary to what might be expected, K is strictly positive. We shall show that a lower bound on K is given by $\delta^2/(1+\delta)^2$ where δ is the positive root of $\delta(2+\delta)^{\lambda-1} = 1$. For $\lambda = 2$, this can be improved to .1925; the same method will yield better values for other choices of λ. For lower density, the following is true.

THEOREM 2. *As $\lambda \to \infty$, $\underline{D}(E_\lambda) = o(\lambda^{-1})$.*

It might be conjectured that this also holds for the upper density, and for the numbers $K = K(\lambda)$.

We first prove that λ^{-1} is an upper bound for $\bar{D}(E_\lambda)$. Consider the integral

$$I = (1/2\pi) \int_0^{2\pi} \{\log M(r) - \log |f(re^{i\theta})|\} d\theta.$$

Let $f(z) = z^p g(z)$, $g(0) \neq 0$. Then, by Jensen's theorem

* Received July 28, 1947.

$$I = \log M(r) - p \log r - (1/2\pi) \int_0^{2\pi} \log | g(re^{i\theta}) | \, d\theta$$
$$\leq \log M(r) - p \log r - \log | g(0) | .$$

Let $H_{r,\lambda}$ be the set of values of θ for which $\log | f(re^{i\theta}) |$ is less than $(1-\lambda) \log M(r)$. By applying to the integral I the identity $\int \phi(x)\, dx = \int_0^\infty \psi(r)\, dr$ where $\phi(x) \geq 0$ and $\psi(r)$ is the measure of the set on which $\phi(x) \geq r$, the integral I may also be expressed as

$$I = (2\pi)^{-1} \log M(r) \int_0^\infty m(H_{r,\lambda})\, d\lambda.$$

Hence, writing $C = \log | g(0) |$, we have

$$(1) \qquad (1/2\pi) \int_0^\infty m(H_{r,\lambda})\, d\lambda \leq 1 - \frac{p \log r + C}{\log M(r)} .$$

Choose R_0 so that $M(r) > 1$ for $r \geq R_0$; then

$$m(z \varepsilon E, R_0 \leq | z | \leq R) = \int_{R_0}^R m(H_{r,\lambda}) r \, dr.$$

Integrating this with respect to λ and using (1), we have

$$(2) \qquad \int_0^\infty D_R(E_\lambda{}^*)\, d\lambda \leqq 1 - R_0{}^2/R^2 - (2/R^2) \int_{R_0}^R \frac{(p \log r + C) r \, dr}{\log M(r)}$$

where $E_\lambda{}^*$ is E_λ with the circle $| z | \leq R_0$ deleted.

We may suppose that $f(z)$ is not a polynomial. (In this case, it is easily seen that $\bar{D}(E_\lambda) = 0$ for all $\lambda > 0$.) Since $\log M(r)$ is convex in $\log r$, it follows that $\log r = o(\log M(r))$ as r tends to infinity, and hence that the right side of (2) is $1 + o(1)$ as $R \to \infty$. As λ increases, the sets $E_\lambda{}^*$ decrease and $D_R(E_\lambda{}^*)$ is monotone for fixed R. Thus, $\lambda D_R(E_\lambda{}^*) \leq \int_0^\infty D_R(E_\lambda{}^*)\, d\lambda$ and letting R increase, we have $\lambda \bar{D}(E_\lambda) = \lambda \bar{D}(E_\lambda{}^*) \leq 1$.

The proof of Theorem 2 also falls out of the inequality (2). Letting R tend to infinity, we have

$$\int_0^\infty \underline{D}(E_\lambda)\, d\lambda \leq 1$$

and since the integrand is monotonic, $\lim_{\lambda \to \infty} \lambda \, \underline{D}(E_\lambda) = 0$.

To obtain a lower bound on K, the least upper bound of $\bar{D}(E_\lambda)$ for all functions $f(z)$, we investigate a special function. Consider the product

$$f(z) = \prod_{n=1}^{\infty} (1 - z/a^n)^{b^n}, \qquad a > b > 1,$$

which defines an entire function of order $\log b / \log a$. Put

$$\phi(z) = |f(z)| M(r)^{\lambda-1} = \prod_{k=1}^{\infty} \{|1 - z/a^k| (1 + r/a^k)^{\lambda-1}\}^{b^k}.$$

Suppose that z lies in the region S described by

$$(3) \qquad |1 - z/a^n| (1 + r/a^n)^{\lambda-1} \leq \beta < 1.$$

Let r/a^n be less than γ for all z in S. Then,

$$\begin{aligned} \phi(z) &\leq \prod_{k<n} (1 + r/a^k)^{\lambda b^k} \beta^{b^n} \prod_{k>n} (1 + r/a^k)^{\lambda b^k} \\ &\leq (\lambda\gamma a^{n-1})^{\lambda b} (\lambda\gamma a^{n-2})^{\lambda b^2} \cdots (\lambda\gamma a)^{\lambda b^{n-1}} \beta^{b^n} \exp\{\lambda\gamma a^n \sum_{k>n} (b/a)^k\} \end{aligned}$$

and

$$\log \phi(z) \leq b^n \left\{ \frac{\lambda \log \lambda\gamma}{b-1} + \frac{\lambda b \log a}{(b-1)^2} + \frac{\lambda b \gamma}{a-b} + \log \beta \right\}.$$

As b and a tend to infinity in such a manner that $b^{-1} \log a$ and b/a approach zero (e. g., $a = b^2$), the bracket approaches $\log \beta$ which is negative. Thus, for any $\beta < 1$ and for suitable a and b, $\phi(z) < 1$ for all z in S, and for the special function that we have constructed, $S \subseteq E_\lambda$.

There is a set of type S enclosing each of the points $z = a^n$. We now estimate the upper density of the union of these sets, and hence the upper density of E_λ. We may take $\beta = 1$. Put $w = z/a^n = \rho e^{i\phi}$; the set S corresponds to the set S^* bounded by the curve $|1 - w|(1 + \rho)^{\lambda-1} = 1$. The circle $|w - 1| < \delta$ where $\delta(2 + \delta)^{\lambda-1} = 1$ lies in S^*. The ratio $D_{1+\delta}(S^*)$ is at least $\delta^2/(1 + \delta)^2$ and since this is independent of n, this number is a lower bound for $K(\lambda)$. A better bound can be obtained by computing the radius ρ_0 for which $D_{\rho_0}(S^*) = m(w \in S^*, |w| \leq \rho_0)/\pi\rho_0^2$ is greatest. This number is then the desired lower bound. In the special case $\lambda = 2$, numerical integration gives the value .1925 for this ratio.

With reference to generalizations, we observe that the relations (1) and (2) hold with $p = 0$ with any subharmonic function $v(z)$ replacing the function $\log |f(z)|$, and with $\mu(r) = \max_\theta v(re^{i\theta})$ replacing $\log M(r)$, provided that $C = v(0)$ is finite. In addition, there is equality instead of inequality in (1) and (2) if $v(z)$ is a harmonic function without singularities.

Brown University.
Society of Fellows, Harvard University.
Syracuse University.

CONCERNING NON-APOSYNDETIC CONTINUA.*

By F. Burton Jones.

Introduction. During the past thirty years considerable attention has been given to the topological properties of compact continua. But until recently very little has been done to fill the gap between indecomposable continua and continuous curves.[1] The work of Whyburn [2] and Wilder [3] on semi-locally-connected continua is a step in this direction. In many ways semi-locally-connected continua strongly resemble continuous curves. So there still exists a gap between *them* and indecomposable continua. This paper is devoted largely to a beginning of the study of the internal properties of continua lying in this gap. These continua are non-aposyndetic and are studied from that point of view mainly because of its simplicity. The other characterizing notion is the weak cut point property. A continuum which is totally non-aposyndetic must contain at least one weak cut point. And such points are everywhere dense in a continuum which is not semi-locally-connected anywhere. It is well known that *every* point of an indecomposable continuum is a weak cut point of it.

1. Definitions and preliminary theorems. Let *space* be metric. A circular region with its center at the point p and its radius equal to $\epsilon > 0$ will be denoted by $U(p, \epsilon)$. A point p of a continuum M is said to be a *weak cut point* of M provided that $M - p$ contains two distinct points x and y such that every subcontinuum of M containing $x + y$ contains p. Under these conditions p is said to *cut x from y in M*. A point p of a continuum M is said to be a *strong cut point* (or a *separating point*) of M provided that $M - p$ is not connected. Under these conditions p is said to *separate x from y in M*. If M is a set, by an *open subset of M* is meant a subset of M

* Received June 10, 1947; Presented to the American Mathematical Society, November 29, 1946, and December 30, 1947.

[1] See R. L. Moore's *Foundations of Point Set Theory*, Karl Menger's *Kurventheorie*, and G. T. Whyburn's *Analytic Topology*.

[2] G. T. Whyburn, "Semi-locally-connected sets," *American Journal of Mathematics*, vol. 61 (1939), pp. 733-749.

[3] R. L. Wilder, "Sets which satisfy certain avoidability conditions," *Casopis pro Pestovani Mathematiky a Fysiky*, vol. 67 (1938), pp. 185-198, and "Property S_n," *American Journal of Mathematics*, vol. 61 (1939), pp. 823-832.

which is open relative to M. A circular open subset of M with center p of M and radius ϵ will also be denoted by $U(p, \epsilon)$ when no confusion results.

Let M be a continuum and let x and y be distinct points of M; if M contains a continuum H and an open subset U such that $M - y \supset H \supset U \supset x$, then M is said to be *aposyndetic* at x with respect to y. Let x be a point of a continuum M; if for each point y of $M - x$, M is aposyndetic at x with respect to y, then M is said to be aposyndetic at x.[4] A continuum M which is *not* aposyndetic at a point x of M is said to be *non-aposyndetic* at x or more briefly, *non-aposyndetic.* A continuum which is non-aposyndetic at each of its points is said to be *totally* non-aposyndetic.

A continuum M is said to be *semi-locally-connected* at a point x of M provided that for each open subset U of M containing x there exists an open subset V of M such that $U \supset V \supset x$ and $M - U$ is contained in the sum of a finite number of components of $M - V$. A continuum which is semi-locally-connected at each of its points is said to be *semi-locally-connected.*[5]

THEOREM 0. *In order that a compact continuum M be semi-locally-connected at a point p of M, it is necessary and sufficient that M be aposyndetic at each point of $M - p$ with respect to p.*

Theorem 0 may be established with the help of the Heine-Borel Theorem.

THEOREM 1. *Suppose that M is a continuum and that $x_1, x_2, x_3, \cdots$ and $y_1, y_2, y_3, \cdots$ are sequences such that* (1) *for each i, x_i is a point of M and y_i is a point of $M - x_i$,* (2) *$x_1, x_2, x_3, \cdots$ converges to a point x and $y_1, y_2, y_3, \cdots$ converges to a point y distinct from x, and* (3) *for each i, M is not aposyndetic at x_i with respect to y_i. Then M is not aposyndetic at x with respect to y.*

Proof. If M were aposyndetic at x with respect to y, M would contain a continuum H and an open subset U such that $M - y \supset H \supset U \supset x$. Hence for some integer i, $M - y_i \supset H \supset U \supset x_i$, and M would be aposyndetic at x_i with respect to y_i contrary to hypothesis.

THEOREM 2. *If x is a point of a continuum M and K is the set of all points y of $M - x$ such that M is not aposyndetic at x with respect to y, then $K + x$ is closed.*

---

[4] F. B. Jones, "Aposyndetic continua and certain boundary problems," *American Journal of Mathematics*, vol. 63 (1941), pp. 545-553.

[5] G. T. Whyburn, *loc. cit.*, p. 734. For the definition of other terms and phrases the reader is referred to Moore's book.

THEOREM 3. *If y is a point of the compact continuum M and L is the set of all points x of $M - y$ such that M is not aposyndetic at x with respect to y, then $L + y$ is a continuum.*

THEOREM 4. *If M is a compact continuum, the set N of points of M at which M is non-aposyndetic is the sum of a countable number of closed sets (i. e., N is an F_σ-set).*

Proof. Let i denote a positive integer and let N_i denote the set of all points x of N such that for some point y of $M - x$, M is not aposyndetic at x with respect to y and $\delta(x, y) \geqq 1/i$. Let α denote a simple sequence of distinct points of N_i converging to a point x_0. For each x of α, let y denote a point of M such that M is not aposyndetic at x with respect to y and $\delta(x, y) \geqq 1/i$, and let β denote the sequence of these points y. Let $y_1, y_2, y_3, \cdots$ denote a subsequence of β converging to a point y_0 of M and let $x_1, x_2, x_3, \cdots$ denote the corresponding subsequence of α. Evidently $x_1, x_2, x_3, \cdots$ converges to the point x_0 and $x_0 \neq y_0$. By Theorem 1, M is not aposyndetic at x_0; hence N_i is closed. But $N = \Sigma N_i$, $i = 1, 2, 3, \cdots$.

THEOREM 5. *If M is a compact continuum, the set of points of M at which M is not semi-locally-connected is the sum of a countable number of closed sets.*

Theorem 5 may be established with the help of Theorem 0 and an argument similar to that used for Theorem 4.

THEOREM 6. *Suppose that* (1) *M is a compact continuum,* (2) *D is an open subset of M, and* (3) *M_0 is a closed subset of M such that $D \cdot M_0 = 0$. Suppose further that for each point x of D, there exists a point y of M_0 such that M is not aposyndetic at x with respect to y. Then if z is a point of $M - M_0$, there exists a point x_0 of D and a point y_0 of M_0 such that y_0 cuts x_0 from z in M.*[6]

Proof. Let z denote a point of $M - M_0$ and let D_1 denote an open subset of M such that $D \supset \bar{D}_1$, $\bar{D}_1 \cdot (M_0 + z) = 0$, and $\delta(\bar{D}_1) < \frac{1}{2}$. Let K_1 denote the set of all points y of M_0 such that for some point x of $\bar{D}_1$, M is not aposyndetic at x with respect to y. The set K_1 is closed and is the sum of a finite collection G_1 of closed subsets each of which has diameter less than $\frac{1}{2}$. For each element g of G_1, let $L(g)$ denote the set of all points x of $\bar{D}_1$ such that for some point y of g, M is not aposyndetic at x with respect to y. For

[6] The hypothesis of Theorem 6 may be weakened in many respects without affecting the validity of the conclusion. For instance, D may have points in common with M_0.

each g of G_1, $L(g)$ is closed and $\bar{D}_1 = \Sigma L(g)$. Hence for some element M_1 of G_1, $L(M_1)$ contains an open subset of D_1. Let V_1 denote an open subset of M containing M_1 such that $\bar{V}_1 \cdot (\bar{D}_1 + z) = 0$ and $\delta(\bar{V}_1) < \frac{1}{2}$. The component C_1 of $M - V_1$ which contains z does not contain an open subset of M containing a point of $L(M_1)$. Hence $L(M_1)$ contains an open subset D_2 of M such that $\bar{D}_2 \cdot C_1 = 0$ and $\delta(\bar{D}_2) < \frac{1}{4}$.

Let K_2 denote the set of all points y of M_1 such that for some point x of $\bar{D}_2$, M is not aposyndetic at x with respect to y. The set K_2 is closed and is the sum of a finite collection G_2 of closed sets each of which has diameter less than $\frac{1}{4}$. For each element g of G_2, let $L(g)$ denote the set of all points x of $\bar{D}_2$ such that for some point y of g, M is not aposyndetic at x with respect to y. For each g of G_2, $L(g)$ is closed and $\bar{D}_2 = \Sigma L(g)$. Hence for some element M_2 of G_2, $L(M_2)$ contains an open subset of D_2. Let V_2 denote an open subset of M containing M_2 such that $V_1 \supset \bar{V}_2$ and $\delta(\bar{V}_2) < \frac{1}{4}$. The component C_2 of $M - V_2$ which contains z does not contain an open subset of M containing a point of $L(M_2)$. Hence $L(M_2)$ contains an open subset D_3 of M such that $\bar{D}_3 \cdot C_2 = 0$ and $\delta(\bar{D}_3) < 1/8$. Continue this process. Let $x_0 = \Pi \bar{D}_i$ and let $y_0 = \Pi M_i = \Pi V_i$. Evidently x_0 is a point of D and y_0 is a point of M_0. Furthermore y_0 cuts x_0 from z in M, for if it did not, then $M - y_0$ would contain a continuum from x_0 to z, and hence for some i, x_0 would have belonged to C_i.

THEOREM 7. *The topological product of a nondegenerate continuum M_1 with a nondegenerate continuum M_2 is aposyndetic at each of its points.*[7]

Proof. Suppose that (a_1, a_2) and (b_1, b_2) are points of $M_1 \times M_2$ such that $M_1 \supset a_1 + b_1$, $M_2 \supset a_2 + b_2$, and $a_1 \neq b_1$. Let (c_1, c_2) denote a point of $M_1 \times M_2$ such that $a_1 \neq c_1 \neq b_1$ and $a_2 \neq c_2 \neq b_2$. Let $M_2(b_1)$ denote the set of all points (x_1, x_2) of $M_1 \times M_2$ such that $x_1 = b_1$. Since $M_2(b_1)$ is a continuum, $M_1 \times M_2$ contains an open set U containing (a_1, a_2) such that $\bar{U} \cdot M_2(b_1) = 0$. Now let $M_1(c_2)$ denote the set of all points (x_1, x_2) of $M_1 \times M_2$ such that $x_2 = c_2$; $M_1(c_2)$ is a continuum not containing (b_1, b_2). Let H_0 be the set of all points (x_1, x_2) of $M_1 \times M_2$ such that for some point y_2 of M_2, (x_1, y_2) belongs to $\bar{U}$. The set H_0 is closed, contains $\bar{U}$, but does not contain (b_1, b_2). Let H denote $H_0 + M_1(c_2)$. Evidently H is a continuum containing U but not (b_1, b_2). Hence $M_1 \times M_2$ is aposyndetic at (a_1, a_2) with respect to (b_1, b_2). It follows that $M_1 \times M_2$ is aposyndetic at each of its points.

---

[7] Cf. G. T. Whyburn, *loc. cit.*, Theorem 4, p. 735.

COROLLARY. *No non-aposyndetic continuum is the topological product of two nondegenerate continua.*

2. Indecomposable continua. An indecomposable continuum M is non-aposyndetic. That such a continuum possesses this property very strongly is shown by Theorems 9 and 10.

THEOREM 8. *In order that a compact continuum M be decomposable it is necessary and sufficient that M contain two points, x and y, such that M is aposyndetic at x with respect to y.*

Proof. The condition is necessary. For suppose that M is the sum of two proper subcontinua H and K. Let x be a point of $M - K$ and y be a point of $M - H$. Then $M - y \supset H \supset M - K \supset x$. Since H is a subcontinuum of M and $M - K$ is an open subset of M, M is aposyndetic at x with respect to y.

The condition is also sufficient; for in this case M contains a continuum H and an open subset U such that $M - y \supset H \supset U \supset x$. If $M - H$ is a connected set V, M is decomposable because $M = H + \bar{V}$. If $M - H$ is not connected, then $M - H$ is the sum of two mutually separate sets A and B. Since H is a continuum, $H + A$ is connected and so is $H + B$. Hence M is decomposable because $M = (H + A) + (H + B)$.

THEOREM 9. *In order that a compact continuum M be indecomposable it is necessary and sufficient that there do not exist two distinct points x and y of M such that M is aposyndetic at x with respect to y.*

THEOREM 10. *In order that a compact continuum M be indecomposable it is necessary and sufficient that M contain an open subset D of M such that if x is a point of D and y is a point of $M - x$, then M is not aposyndetic at x with respect to y.*

Proof. The necessity of the condition follows from Theorem 9 by letting D be M. The condition is also sufficient; for suppose that M may be decomposed into the continua H and K. Obviously M is aposyndetic at each point of $M - K$ with respect to a point (any point) of $M - H$; and conversely. Hence $D \cdot [(M - H) + (M - K)] = 0$. Therefore D is a subset of $H \cdot K$. Let x be a point of D and y a point of $M - H$. Then $M - y \supset H \supset D \supset x$ and M is aposyndetic at x with respect to y. This is a contradiction.

Example 0. Let H and K denote compact, plane, indecomposable

continua whose common part is an arc T and let $M = H + K$. If x is a point of T and y is a point of $M - x$, M is not aposyndetic at x with respect to y. But M is decomposable. Hence in the hypothesis of Theorem 10 the stipulation that D be an open subset of M may not be replaced by the stipulation that D be a continuum.

3. Fundamental cut point existence theorems. Continua which are totally non-aposyndetic or approximately so must contain one or more weak cut points. The more nearly a continuum approximates an indecomposable continuum in its non-aposyndetic properties, the more nearly it approximates the cut point properties of one.

THEOREM 11. *If z is a point of a compact continuum M and D is an open subset of $M - z$ such that* (1) *M is non-aposyndetic at every point of D but* (2) *M is aposyndetic at each point of D with respect to z, then D contains a point x and M contains a point y such that y cuts x from z in M.*

Proof. There exists an open subset D_1 of M such that $D \supset \bar{D}_1$. Let ϵ denote a positive number and let $K(\epsilon)$ denote the set of all points x of $\bar{D}_1$ such that for some point y of M, M is not aposyndetic at x with respect to y and $\delta(x, y) \geqq \epsilon$. Let α denote a sequence of points of $K(\epsilon)$ converging to a point $\bar{x}$. There exist a subsequence $x_1, x_2, x_3, \cdots$ of α and a sequence $y_1, y_2, y_3, \cdots$ of points of M converging to a point $\bar{y}$ of M such that for each i, $\delta(x_i, y_i) \geqq \epsilon$ but M is not aposyndetic at x_i with respect to y_i. Obviously $\delta(\bar{x}, \bar{y}) \geqq \epsilon$ and $\bar{x} \neq \bar{y}$. It follows from Theorem 1 that $\bar{x}$ belongs to $K(\epsilon)$. Hence $K(\epsilon)$ is closed (if it exists). Since for each x of $\bar{D}_1$, there is some ϵ such that $K(\epsilon)$ contains x, it follows that $\bar{D}_1 = \Sigma K(1/n)$, $n = 1, 2, 3, \cdots$. So for some ϵ, $K(\epsilon)$ contains an open subset of M and, in particular, contains an open subset D_2 of M such that $\delta(\bar{D}_2) < \frac{1}{4}\epsilon$. Let M_0 denote the set of all points y of M such that for some point x of $\bar{D}_2$, $\delta(x, y) \geqq \epsilon$ but M is not aposyndetic at x with respect to y. Obviously $D_2 \cdot M_0 = 0$, and z is a point of $M - M_0$. Hence, by Theorem 6, there exist a point x of D_2 and a point y of M_0 such that y cuts x from z in M.

THEOREM 12. *If D is an open subset of a compact continuum M such that M is non-aposyndetic at every point of D, then D contains a point x and M contains points y and z such that y cuts x from z in M.*

Proof. In case M is indecomposable the conclusion to Theorem 12 easily follows since (1) each composant of M is dense in M and (2) any point y of M cuts x from z in M if x and z are points of different composants of M.

On the other hand, if M is decomposable, then, by Theorem 10, D contains a point p and $M - p$ contains a point z such that M is aposyndetic at p with respect to z. Hence M contains a continuum H and an open subset U such that $M - z \supset H \supset U \supset p$. Let $D' = D \cdot U$. It follows from Theorem 11 that D' contains a point x and M contains a point y such that y cuts x from z in M.

COROLLARY. *A totally non-aposyndetic compact continuum contains at least one weak cut point.*

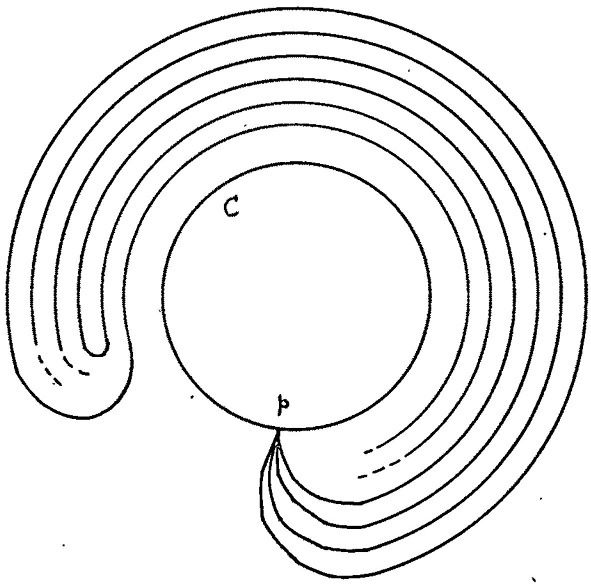


Figure 1.

Example 1. While a totally non-aposyndetic compact continuum must contain at least one weak cut point, it need *not* contain *more* than one. To see this let M be a plane continuum composed of a circle C together with a bundle of rays (emanating from a point p on C), whose cross-section is a Cantor discontinuum and which is wrapped back and forth (as indicated in Figure 1) so that each ray has C as its limiting set. This continuum has no strong cut point at all and only one weak cut point—namely, p. But M is non-aposyndetic at each point of $M - p$ with respect to p, and M is non-aposyndetic at p with respect to each point of $C - p$.

THEOREM 13. *Suppose that* (1) D *is an open subset of a compact continuum* M *and* (2) *for each point* p *of* D, *there exists a point* o *of* $M-p$ *such that* M *is not aposyndetic at* o *with respect to* p. *Then* D *contains distinct points* x, y, *and* z, *such that* y *cuts* x *from* z *in* M.

Proof. Let D_1 denote an open subset of M such that $D \supset \bar{D}_1$. For each positive integer n, let M_n denote the set of all points p of $\bar{D}_1$ such that for some point q of M, M is not aposyndetic at q with respect to p and $\delta(p,q) \geq 1/n$. Since $\bar{D}_1 = \Sigma M_n$ and, for each n, M_n is closed, it follows that, for some n, M_n contains an open subset of M. Hence there exist a point A and a positive number ϵ such that for each point p of the circular open subset $U(A,\epsilon)$ of M (and of M_n) there exists a point q of $M-U(A,\epsilon)$ such that M is not aposyndetic at q with respect to p. Let $U(A,\epsilon/4)$ and $U(A,\epsilon/2-\epsilon/4n)$ denote circular open subsets of M, and let $R_1 = U(A,\epsilon/4)$ and $R_n = U(A,\epsilon/2-\epsilon/4n)-\bar{R}_{n-1}$, $n \geq 2$. Let $R_\infty = U(A,\epsilon/2)$.

Now assume Theorem 13 to be false. It follows from this assumption and Theorem 6 that R_1 contains a point x_1 such that if y_1 is a point of $\bar{R}_\infty - R_1$, M is aposyndetic at x_1 with respect to y_1. There exist a finite sequence $H_{11}, H_{12}, \cdots, H_{1i}, \cdots, H_{1n_1}$ of continua of M and a finite sequence $U_{11}, U_{12}, \cdots, U_{1n_1}$ of open subsets of M such that for each point y_1 of $\bar{R}_\infty - R_1$, there exists an integer j $(1 \leq j \leq n_1)$, so that $M - y_1 \supset H_{1j} \supset U_{1j} \supset x_1$. Let V_1 denote a circular open subset of M (of radius less than $\frac{1}{2}\epsilon$) such that $R_1 \cdot \Pi U_{1i} \supset V_1 \supset x_1$. Since there is a point of $M-U(A,\epsilon)$ at which M is not aposyndetic with respect to x_1, $M - V_1$ is the sum of two mutually exclusive closed sets Q_1 and W_1 each of which contains a point of $M-U(A,\epsilon)$. Hence each of the sets W_1 and Q_1 contains a point of R_2. Again it follows (from assuming that Theorem 13 is false) that there exists a point x_2 of $R_2 \cdot W_1$ such that for each point y_2 of $\bar{R}_\infty - (\bar{R}_1 + R_2)$, M is aposyndetic at x_2 with respect to y_2. There exist a finite sequence $H_{21}, H_{22}, \cdots, H_{2i}, \cdots, H_{2n_2}$ of continua of M and a finite sequence $U_{21}, U_{22}, \cdots, U_{2n_2}$ of open subsets of M such that for each point y_2 of $\bar{R}_\infty - (\bar{R}_1 + R_2)$, there exists an integer j $(1 \leq j \leq n_2)$, so that $M - y_2 \supset H_{2j} \supset U_{2j} \supset x_2$. Furthermore there exists a continuum H_1 such that (1) for some j, $H_1 = H_{1j}$, and (2) H_1 does not contain x_2. Let V_2 denote a circular open subset of M (of radius less than $\frac{1}{4}\epsilon$) such that $(M-H_1) \cdot R_2 \cdot \Pi U_{2i} \supset V_2 \supset x_2$. Since there is a point of $M-U(A,\epsilon)$ at which M is not aposyndetic with respect to x_2, $M - V_2$ is the sum of two mutually exclusive closed sets Q_2 and W_2 such that (1) Q_2 contains x_1 and (2) each of the sets Q_2 and W_2 contains a point of $M-U(A,\epsilon)$. Hence each of the sets W_2

and Q_2 contains a point of R_3. Furthermore $Q_2 \supset H_1 \supset V_1$ and consequently Q_2 contains Q_1 and W_1 contains W_2.

This process may be continued. Hence there exist a point y of the boundary (with respect to M) of R_∞ and infinite sequences (these are possibly subsequences of those chosen in the above process): $x_1, x_2, x_3, \cdots$ of points of M, $V_1, V_2, V_3, \cdots$ of open subsets of M, and $H_1, H_2, H_3, \cdots$ of subcontinua of M such that (1) $x_1, x_2, \cdots$ and $V_1, V_2, \cdots$ converge to y, (2) for each i, $R_\infty \supset V_i$, (3) for each i, $M - V_{i+1} \supset H_i \supset V_i \supset x_i$, and (4) for each i, $M - V_i$ is the sum of two mutually exclusive closed sets W_i and Q_i so that each contains a point of $M - U(A, \epsilon)$ and W_i contains V_{i+1} but not x_{i-1}. It follows from (3) and (4) that for each i $(i \geq 2)$, W_i lies in W_{i-1} and contains a point of β, the (relative) boundary of $U(A, \epsilon)$. Let z denote a point of $\Pi\beta \cdot W_i$. Evidently for each i $(i \geq 2)$, V_i separates x_1 from z in M. It follows from (1) that y cuts x_1 from z in M.

THEOREM 14. *Suppose that M is a compact continuum such that for each point p of M, there exists a point q of M such that M is not aposyndetic at q with respect to p. Then the set of weak cut points of M is dense in M.*

Theorem 14 follows immediately from Theorem 13. The reader should compare Theorem 14 with the Corollary to Theorem 12. To emphasize the differences between Theorems 12 and 13, the following theorem may be established with the help of Theorems 0 and 13.

THEOREM 15. *If a compact continuum M is not semi-locally-connected at any point of an open subset U of M, then the set of weak cut points of M is dense in U.*

COROLLARY. *If a compact continuum is not semi-locally-connected at any of its points, then the set of its weak cut points is dense in it.*

4. Totally non-aposyndetic continua which contain only one weak cut point. The example following Theorem 12 shows that continua exist (even in the plane) which are totally non-aposyndetic but contain only *one* weak cut point. While this example appears special, it is nevertheless general in some of its properties. Two of these properties are pointed out in the following two theorems and a third is contained in Theorem 15.

THEOREM 16. *If the nondegenerate continuum M has only one weak cut point, the composant of M containing P is M.*

Proof. Let x denote a point of M distinct from p, and let y denote a point of $M - (p + x)$. Since y does not cut x from p in M, $M - y$ contains a continuum containing $p + x$.

THEOREM 17. *If the compact continuum M is totally non-aposyndetic but has only one weak cut point p, then M is non-aposyndetic at each point of $M - p$ with respect to p.*

Proof. Suppose, on the contrary, that M is aposyndetic at a point x of $M - p$ with respect to p. Then M contains a continuum H and an open subset U such that $M - p \supset H \supset U \supset x$. Evidently M is aposyndetic at every point of U with respect to p. It follows from Theorem 11 that some point y of $M - p$ cuts p from a point of U. This is a contradiction.

COROLLARY. *If the compact continuum M is totally non-aposyndetic but has only one weak cut point p, then M is not semi-locally-connected at p.*

Example 2. Let p be a point in a Euclidean 3-space. Let G be a collection of circles all of equal radii and perpendicular to the same plane E such that (1) the intersection of any two elements of G is p, (2) E contains p, and (3) $E \cdot G^* - p$ is a Cantor discontinuum (G^* is the sum of the elements of G). Now let $M = G^*$. The continuum M is compact and has p as its only cut point. Furthermore M fails to be semi-locally-connected only at the point p. But M is aposyndetic at no point except p. In Example 1 (where M is non-aposyndetic at *every* point) M fails to be semi-locally-connected at some point other than its one cut point p. However, the reader will notice considerable similarity between Examples 1 and 2. The author feels that this similarity is not accidental but is a consequence of some general decomposition theorem which is as yet not known.

5. Continua containing no weak cut point. Two fundamental theorems concerning the existence of points in a continuum at which a continuum is aposyndetic or at which a continuum is semi-locally-connected follow almost at once from the preceding theorems. These will be stated with only a reference to the theorems helpful in their proof.

THEOREM 18. *If a compact continuum M contains no weak cut point, then the set of points at which M is both aposyndetic and semi-locally-connected (simultaneously) is a dense inner-limiting set (i. e., a G_δ-set).*[8]

---

[8] Theorem 18 is a sort of converse to Theorem 6.21 of Whyburn's "Semi-locally-connected sets," *loc. cit.* In connection with his paper, "Some characterizations of

Theorem 18 may be proved with the help of Theorems 4, 5, 12, and 15.

THEOREM 19. *If the open subset U of the compact continuum M contains no weak cut point of M, then U contains a point at which M is semi-locally-connected.*

Theorem 19 follows from Theorem 15.

Of course, a continuum which contains no weak cut point need not be aposyndetic although from a certain point of view it is nearly so.

Example 3. Let M be a continuum in the Euclidean number plane consisting of: (1) a rectangle R whose corners are $(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$, (2) those vertical intervals joining a point of the Cantor discontinuum

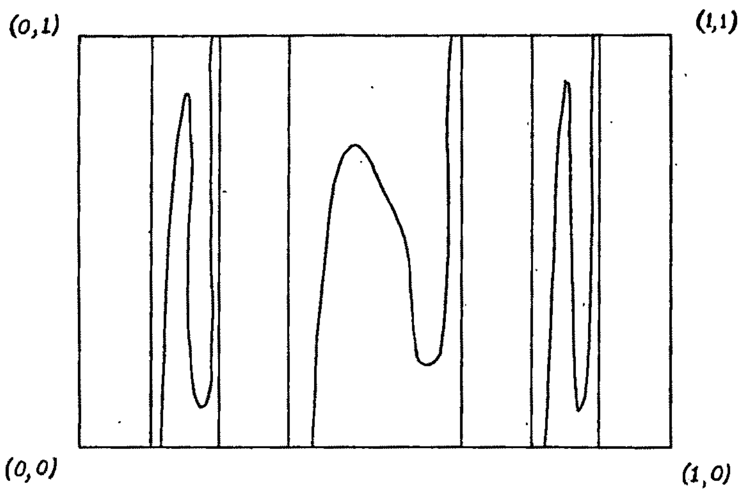


Figure 2.

in the base of R to a point on the opposite side, and (3) infinitely many S-curves joining the base of R to the opposite side as indicated in Figure 2. The continuum M is both aposyndetic and semi-locally-connected at each point of an S-curve and at each point of the top and bottom of R not in its Cantor set but nowhere else. Obviously, M contains no weak cut point.

THE UNIVERSITY OF TEXAS.

---

simple closed curves," *American Journal of Mathematics*, vol. 70 (1948), pp. 497-506, R. H. Bing raised with me the question: *Does a compact continuum containing no weak cut point have to contain a point at which it is aposyndetic?* Theorem 18 answers this question in the affirmative.

THE CONFIGURATION OF SCHUR QUADRICS AND THE PARABOLIC CURVE OF THE TRINODAL CUBIC SURFACE.*

By S. BISHARA and A. Y. AMIN.

1. Introduction. It is a known property of the general cubic surface that corresponding to every configuration of 12 lines of the surface forming a double-six as e. g.,

$$\begin{matrix} a_1a_2a_3a_4a_5a_6 \\ b_1b_2b_3b_4b_5b_6 \end{matrix}$$

there exists a unique quadric called the "Schur quadric" of the double-six with respect to which corresponding lines of the double-six are polar reciprocals, i. e., it reciprocates a_1 into b_1, a_2 into b_2, etc. This quadric passes through the 24 parabolic points of the 12 lines of the double-six and it reciprocates the cubic surface containing these 12 lines into the cubic envelope containing the same lines. A Schur quadric depends for its determination only on three pairs of corresponding lines. In fact there exists a unique quadric which reciprocates a skew hexagon into itself i. e., reciprocates every edge into the opposite edge.

Since the general cubic surface contains 36 double-sixes it follows that there are 36 Schur quadrics for the general cubic. These quadrics play an important rôle in the theory of the cubic surface. They group themselves:

I. Either into triads corresponding to three associated double-sixes as for example those indicated by:

$$\begin{matrix} a_2b_2c_{31}c_{34}c_{35}c_{36} & a_1b_1c_{32}c_{34}c_{35}c_{36} & a_1b_1c_{23}c_{24}c_{25}c_{26} \\ a_3b_3c_{21}c_{24}c_{25}c_{26} & a_3b_3c_{12}c_{14}c_{15}c_{16} & a_2b_2c_{13}c_{14}c_{15}c_{16} \end{matrix}$$

in which every line occurs in two of them and these quadrics have the following properties: [1]

(i) They satisfy the invariant relation $\Theta\Theta' = 4\Delta\Delta'$ for any pair.

(ii) The three quadrics are inscribed in a common developable, i. e., their tangential equations are linearly connected.

* Received May 12, 1947.

[1] A. L. Dixon, *Journal of the London Mathematical Society*, vol. 1 (1926), pp. 170-175 and *Proceedings of the London Mathematical Society* (2), vol. 26 (1926), pp. 351-362

(iii) The Hessian can be expressed in the form:

$$H - \psi\pi = S_2S_3 + S_3S_1 + S_1S_2$$

where S_1, S_2, S_3 are the three Schur quadrics, H is the Hessian, ψ the cubic surface and π the polar plane with respect to the cubic surface of F_{123},[2] the point through which pass all planes cutting the cubic surface in cubic curves which are met by a_1, a_2, a_3 in three points forming an apolar triad on the plane cubic of section.

II. Or in pairs corresponding to double-sixes of the type:

$$\begin{matrix} a_1a_2a_3a_4a_5a_6 & \quad a_1b_1c_{23}c_{24}c_{25}c_{26} \\ b_1b_2b_3b_4b_5b_6 & \quad a_2b_2c_{13}c_{14}c_{15}c_{16} \end{matrix}$$

in which two lines of the upper row (or lower row) of the first are corresponding lines of the second and these have the characteristic property that they satisfy the invariant relations $\Theta = 0$, $\Theta' = 0$.[3]

One of the main objects of the present paper is to investigate the configuration of Schur quadrics for the trinodal cubic surface, their mutual relationships and their characteristic properties.

2. The Schur quadrics of the trinodal cubic surface. Let the three nodes of the trinodal cubic surface be taken for vertices A, B, C of the tetrahedron of reference and let the fourth vertex D be the point in which the three stationary tangent planes along BC, CA, AB intersect. The equation of the trinodal cubic surface can then be put in the form:

$$\psi \equiv dt^3 + 3d_1t^2x + 3d_2t^2y + 3d_3t^2z + 6sxyz = 0.$$

The 27 lines of the general surface reduce for the trinodal cubic into 12 owing to the following coincidences:

$$a_1 \equiv a_2; \quad a_3 \equiv a_4; \quad a_5 \equiv a_6; \quad b_1 \equiv b_2; \quad b_3 \equiv b_4; \quad b_5 \equiv b_6;$$
$$c_{35} \equiv c_{45} \equiv c_{36} \equiv c_{46}; \quad c_{15} \equiv c_{16} \equiv c_{25} \equiv c_{26}; \quad c_{13} \equiv c_{14} \equiv c_{23} \equiv c_{24}.$$

[2] W. P. Milne, *Proceedings of the London Mathematical Society* (2), vol. 21 (1921), p. 134.

[3] S. Bishara, *Journal of the London Mathematical Society*, vol. 6, Part I (1931), p. 12.

The equations of the 12 distinct lines of the surface are as follows:

Line	Equations
$c_{35} \equiv c_{45} \equiv c_{36} \equiv c_{46}$	$x = 0, \quad t = 0$
$c_{15} \equiv c_{16} \equiv c_{25} \equiv c_{26}$	$y = 0, \quad t = 0$
$c_{13} \equiv c_{14} \equiv c_{23} \equiv c_{24}$	$z = 0, \quad t = 0$
c_{12}	$x = 0, \quad dt + 3d_2y + 3d_3z = 0$
c_{34}	$y = 0, \quad dt + 3d_1x + 3d_3z = 0$
c_{56}	$z = 0, \quad dt + 3d_1x + 3d_2y = 0$
$a_1 \equiv a_2$	$dt + 3d_2y + 3d_3z = 0, \quad t = 3d_3k_1z$
$b_1 \equiv b_2$	$dt + 3d_2y + 3d_3z = 0, \quad t = 3d_3k_2z$
$a_3 \equiv a_4$	$dt + 3d_1x + 3d_3z = 0, \quad t = 3d_1k_1x$
$b_3 \equiv b_4$	$dt + 3d_1x + 3d_3z = 0, \quad t = 3d_1k_2x$
$a_5 \equiv a_6$	$dt + 3d_1x + 3d_2y = 0, \quad t = 3d_2k_1y$
$b_5 \equiv b_6$	$dt + 3d_1x + 3d_2y = 0, \quad t = 3d_2k_2y$

where k_1 and k_2 are the roots of the quadratic:

$$9d_1d_2d_3k^2 - 2dsk - 2s = 0$$

$$k_1 = \{ds + (d^2s^2 + 18sd_1d_2d_3)^{1/2}\}/9d_1d_2d_3$$

and

$$k_2 = \{ds - (d^2s^2 + 18sd_1d_2d_3)^{1/2}\}/9d_1d_2d_3.$$

The 36 double-sixes of the general cubic surface reduce for the trinodal cubic into 14 distinct double-sixes owing to the coincidences of the lines referred to. The equations of the Schur quadrics for these double-sixes are as follows:

The standard double-six:

$$\begin{matrix} a_1a_2a_3a_4a_5a_6 \\ b_1b_2b_3b_4b_5b_6 . \end{matrix}$$

Its Schur quadric is:

$$G \equiv (d^2s - 9d_1d_2d_3)t^2 + 6s(3d_2d_3yz + 3d_3d_1zx + 3d_1d_2xy + dd_1xt + dd_2yt + dd_3zt) = 0.$$

The group of 15 double-sixes obtained by taking any two lines of the upper row (or lower row) of the standard double-six for corresponding lines:

The double-six:

$$\begin{matrix} a_1b_1c_{23}c_{24}c_{25}c_{26} \\ a_2b_2c_{13}c_{14}c_{15}c_{16}. \end{matrix}$$

Its Schur quadric is the nodal cone at A namely:

$$d_1t^2 + 2syz = 0.$$

The double-six (a_3, a_4) has for its Schur quadric the nodal cone at B namely

$$d_2t^2 + 2szx = 0.$$

The double-six (a_5, a_6) has for its Schur quadric the nodal cone at C namely

$$d_3t^2 + 2sxy = 0.$$

The double-six (a_1, a_3) has for its Schur quadric the plane pair:

$$t(dt + 6d_3z) = 0.$$

The last double-six represents in fact 4 coincident double-sixes defined by:

$$(a_1, a_3)\ ; \quad (a_1, a_4)\ ; \quad (a_2, a_3)\ ; \quad (a_2, a_4).$$

Similarly the double-six (a_3, a_5) has for its Schur quadric the plane pair:

$$t(dt + 6d_1x) = 0$$

and it represents 4 coincident double-sixes defined by:

$$(a_3, a_5)\ ; \quad (a_3, a_6)\ ; \quad (a_4, a_5)\ ; \quad (a_4, a_6).$$

Finally the double-six (a_1, a_5) has for its Schur quadric the plane pair:

$$t(dt + 6d_2y) = 0$$

and it amounts to four coincident double-sixes namely:

$$(a_1, a_5)\ ; \quad (a_1, a_6)\ ; \quad (a_2, a_5)\ ; \quad (a_2, a_6).$$

The group of 20.

The Schur quadric for the double-six (a_1, c_{23}) or (a_1, c_{24}) is a cone with vertex B containing the lines BC, BA given by:

$$dt^2 - 18d_1d_3k_1zx + 6d_1xt + 6d_3zt = 0.$$

and the Schur quadric for the double-six (b_1, c_{23}) or (b_1, c_{24}) is a cone with vertex B, also containing the lines BC, BA given by:

$$dt^2 - 18d_1d_3k_2zx + 6d_1xt + 6d_3zt = 0.$$

Similarly for the other two pairs of Schur cones whose vertices are at A and C.

Finally the double-six (a_1, c_{35}) has for its Schur quadric the plane of the nodes taken twice over i. e. $t^2 = 0$. This double-six amounts to 8 coincident double-sixes namely:

$$(a_1, c_{35})\,; \quad (a_1, c_{36})\,; \quad (a_1, c_{45})\,; \quad (a_1, c_{46})\,;$$
$$(b_1, c_{35})\,; \quad (b_1, c_{36})\,; \quad (b_1, c_{45})\,; \quad (b_1, c_{46}).$$

The following results are obtained:

I. The genuine Schur quadric G of the standard double-six (a_1, b_1) can be expressed in the alternative form:

$$9(d_1^2sx^2 + d_2^2sy^2 + d_3^2sz^2 + d_1d_2d_3t^2) - s(3d_1x + 3d_2y + 3d_3z + dt)^2 = 0$$

i. e. as the sum of five squares showing that the five faces of the Steiner-trihedral pair consisting of:

(i) the three stationary tangent planes of the cubic surface,

(ii) the plane of the nodes together with the tritangent plane containing c_{12}, c_{34}, c_{56}

form a self-conjugate pentahedron for the quadric Schur surface.

II. It is known that if two lines l_1 and l_2 intersecting at O are polar lines with respect to a given quadric S then S passes through O touching the plane of l_1 and l_2, the generators of S through O being harmonic conjugates with respect to l_1 and l_2.

Again if the vertex of a cone lies on a given quadric then if the cone regarded as a locus is apolar to the quadric regarded as an envelope, the two generators of the quadric at the common point harmonically separate the two generators of the cone lying in the tangent plane to the quadric at the common point.

For let the quadric be:

(1) $$xt = yz\,;$$

its tangential equation is:

(2) $$lp = mn.$$

Let the cone be:

(3) $$ax^2 + by^2 + cz^2 + 2fyz + 2gzx + 2hxy = 0$$

whose vertex is D. (2) and (3) are apolar if

$$(4) \qquad f = 0.$$

Hence DB, DC are conjugate lines with respect to the two generators of the cone in the plane CDB.

From the above results it follows that the two generators of G at the node A harmonically separate a_1 and b_1 which are the lines of intersection of the nodal cone at A with the tangent plane to G at A. Hence the nodal cone at A as a locus is apolar to G as an envelope. Also the nodal cone at A as an envelope is apolar to G as a locus since G passes through A.

Hence

$$\Theta = 0, \qquad \Theta' = 0.$$

This is in conformity with the result mentioned in **1.**

III. The two Schur cones of the double-sixes (a_1, c_{23}) and (b_1, c_{23}) namely:

$$dt^2 - 18d_1d_3k_1zx + 6d_1xt + 6d_3zt = 0$$

and

$$dt^2 - 18d_1d_3k_2zx + 6d_1xt + 6d_3zt = 0$$

have in common the four generators:

$$x = 0, \quad t = 0; \qquad z = 0, \quad t = 0; \qquad x = 0, \quad dt + 6d_3z = 0;$$
$$z = 0, \quad dt + 6d_1x = 0.$$

But the Schur quadric of the double-six (a_3, a_5) is the plane pair $t(dt + 6d_1x) = 0$ and that of (a_1, a_3) is $t(dt + 6d_3z) = 0$.

Hence the two Schur cones at B (other than the nodal cone at B) intersect in four straight lines namely the lines of intersection of:

(i) the Schur plane pair whose edge is AB with the plane DBC.

(ii) the Schur plane pair whose edge is BC with the plane DAB.

IV. The Schur quadric of the double-six $(a_1; a_3)$ is the plane pair $t(dt + 6d_3z) = 0$. The two planes $t = 0$, $dt + 6d_3z = 0$ harmonically separate the two planes a_1b_3 and a_3b_1. But from the apolarity of this Schur quadric with the quadric G, it follows that the two Schur planes through AB harmonically separate the two tangent planes drawn from AB to the quadric G. Hence the two Schur planes through AB are the two planes

which harmonically separate the pair of planes a_1b_3 and a_3b_1 and the pair of tangent planes from AB to the quadric G.

3. The locus of parabolic curves of a pencil of trinodal cubic surfaces. The locus of parabolic curves of a pencil of cubic surfaces of general character through the nine lines of intersection of a pair of conjugate Steinerian trihedra is a desmic quartic surface[4] possessing 12 nodes. These nodes fall in three sets of tetrads of points which are quadruply in perspective i. e., every straight line joining a node of the first set to a node of the second must pass through a node of the third.

We consider in this connection the locus of parabolic curves of the pencil of trinodal cubic surfaces through the nine lines:

$$\begin{array}{ccc} c_{12} & c_{14} & c_{16} \\ c_{32} & c_{34} & c_{36} \\ c_{52} & c_{54} & c_{56}. \end{array}$$

The pencil of trinodal cubic surfaces through these nine lines is given by

$$\psi \equiv dt^3 + 3d_1t^2x + 3d_2t^2y + 3d_3t^2z + 6sxyz = 0.$$

The various members of the pencil are obtained by assigning arbitrary values to the parameter s. The Hessians of the various members are given by

$$H \equiv t^2(d_1{}^2x^2 + d_2{}^2y^2 + d_3{}^2z^2 - 2d_2d_3yz - 2d_3d_1zx - 2d_1d_2xy) + 2sxyz(d_1x + d_2y + d_3z + dt) = 0.$$

The locus of parabolic curves of the pencil is obtained by eliminating s between H and ψ. This is a quartic surface Γ given by:

$$\Gamma \equiv t^2(d^2t^2 + 12d_2d_3yz + 12d_3d_1zx + 12d_1d_2xy + 4dd_1xt + 4dd_2yt + 4dd_3zt) = 0$$

consisting of the plane of the nodes taken twice over and the quadric P where:

$$P \equiv d^2t^2 + 12d_2d_3yz + 12d_3d_1zx + 12d_1d_2xy + 4dd_1xt + 4dd_2yt + 4dd_3zt = 0.$$

The quartic Γ can also be expressed in one of the following forms:

(i) $$\Gamma \equiv t(dt + 6d_2y)\cdot t(dt + 6d_3z) + t(dt + 6d_3z)\cdot t(dt + 6d_1x) + t(dt + 6d_1x)\cdot t(dt + 6d_2y) = 0$$

[4] S. Bishara, *Proceedings of the London Mathematical Society*, Ser. 2, vol. 35 (1933), p. 241.

i. e., it is expressible in terms of the Schur quadrics of the double-sixes defined by:

$$(a_1, a_3); \qquad (a_3, a_5); \qquad (a_5, a_1).$$

(ii) $$\Gamma \equiv t^2(\lambda t^2 + G) = 0$$

i. e., it is also expressible in terms of the Schur quadrics of the double-sixes (a_1, b_1); (b_1, c_{35}); (a_1, c_{35}). From this we obtain the following results:

I. The quadric P is a cone whose vertex L is the point of intersection of the three planes: $dt + 6d_1x = 0$, $dt + 6d_2y = 0$, $dt + 6d_3z = 0$

i. e.

$$L \equiv (1/d_1, 1/d_2, 1/d_3, -6/d).$$

II. The quadric cone P has ring contact with the quadric G, the plane of contact being the plane of nodes $t = 0$.

III. The polar plane of the vertex L of the cone P with respect to the trinodal cubic surface is the plane $d_1x + d_2y + d_3z = 0$, i. e. the plane joining D to the line of intersection of the plane of c_{12}, c_{34}, c_{56} with the plane ABC.

IV. The plane joining the node A to the line c_{12} i. e. (A, c_{12}) is $3d_2y + 3d_3z + dt = 0$ which goes through L. Similarly the planes:

$$(B, c_{34}) \equiv 3d_3z + 3d_1x + dt = 0 \text{ and } (C, c_{56}) \equiv 3d_1x + 3d_2y + dt = 0$$

go through L. Hence L is the point of intersection of the three planes:

$$(A, c_{12}); \qquad (B, c_{34}); \qquad (C, c_{56}).$$

V. The harmonic conjugates of the plane ABC with respect to the pairs of planes (a_1, b_3), (a_3, b_1); (a_3, b_5), (a_5, b_3); (a_1, b_5), (a_5, b_1) all pass through L i. e. the Schur quadrics of the double-sixes (a_1, a_3), (a_3, a_5), (a_5, a_1) all pass through L.

VI. The vertex L of the cone P possesses also the following characteristic property:

Of the pencil of trinodal cubic surfaces through the nine lines:

$$\begin{matrix} c_{12} & c_{14} & c_{16} \\ c_{32} & c_{34} & c_{36} \\ c_{52} & c_{54} & c_{56} \end{matrix}$$

there is only one member having a fourth node (i. e. a quadri-nodal cubic surface). The fourth node is the vertex L of the cone P. The nodal cone of the quadri-nodal cubic member at the fourth node (other than A, B, C) is the cone P.

We conclude with the following results:

I. The twelve lines of the trinodal cubic surface are bitangents to $\Gamma \equiv t^2P = 0$. Three of the lines lie in the plane of the nodes. The nine remaining lines are tangents to P. P goes through A, B, C. The tangent plane at A to P is the plane (a_1b_2). The tangent plane at B is the plane (a_3b_4). The tangent plane at C is the plane (a_5b_6). The vertex L is therefore the point in which the planes (a_1b_2), (a_3b_4), (a_5b_6) intersect. These three planes are however identical with the planes (A, c_{12}), (B, c_{34}), (C, c_{56}) since c_{12} lies in the plane (a_1b_2) etc.

The lines c_{12}, c_{34}, c_{56} are tangents to the cone P. Its point of contact with c_{12} is the point $(0, 1/d_2, 1/d_3, -6/d)$ which lies on AL.

II. The quartic surface Γ can be expressed in the form:

$$\Gamma \equiv \psi(d_1x + d_2y + d_3z + dt) - 3H.$$

The plane $d_1x + d_2y + d_3z + dt = 0$ is identified as the polar plane with respect to the cubic surface of the point D in which the three stationary tangent planes along BC, CA, AB intersect. D is the point F_{135} i. e., the point through which pass all planes cutting the cubic surface in cubic curves which are met by a_1, a_3, a_5 in three points forming an apolar triad on the plane cubic of section. The polar plane of the point $F(a_1, b_1, c_{35})$ with respect to the trinodal cubic surface is the same as the polar plane of the point $F(a_1, a_3, a_5)$.

FAROUK I UNIVERSITY,
ALEXANDRIA, EGYPT.

ON THE IRREDUCIBILITY OF POLYNOMIALS WITH LARGE THIRD COEFFICIENT.*

By Alfred Brauer.

Introduction. Perron [1] was the first who obtained criteria for the irreducibility of polynomials depending on the comparative size of the coefficients. In particular, he proved that a polynomial

$$(1) \qquad f(x) = x^n + a_1 x^{n-1} + \cdots + a_n \qquad (a_n \neq 0)$$

with integral rational coefficients is irreducible in the field of rational numbers P if

$$| a_1 | > 1 + | a_2 | + \cdots + | a_n | .$$

It follows from a theorem of Berwald [2] that the weaker condition

$$| a_1 | > | 1 + a_2 | + | a_3 | + \cdots + | a_n |$$

is sufficient for irreducibility in P. This was noticed by Lipka.[3] Nagell [4] proved that (1) is irreducible in P if

$$| a_{n-1} | > 1 + | a_{n-2} a_n | + | a_{n-3} a_n{}^2 | + \cdots + | a_1 a_n{}^{n-2} | + | a_n{}^{n-1} | .$$

In the following, we consider polynomials with large third coefficient. Perron [5] obtained the following results for such polynomials.

Theorem I. *Set* $A = | a_1 | + | a_2 | + \cdots + | a_n |$ *and*

$$P = | a_1/i\sqrt{a_2} + a_3/(i\sqrt{a_2})^3 + \cdots + a_n/(i\sqrt{a_2})^n | .$$

* Received July 17, 1947.

[1] O. Perron, "Neue Kriterien für die Irreduzibilität algebraischer Gleichungen," *Journal für die reine und angewandte Mathematik*, vol. 132 (1907), pp. 288-307. Cf. O. Perron, *Algebra*, vol. II, 2nd ed. Berlin, 1933, pp. 34-35.

[2] L. Berwald, "Über einige mit dem Satz von Kakeya verwandte Sätze," *Mathematische Zeitschrift*, vol. 37 (1933), pp. 61-76.

[3] St. Lipka, "Többtagiak Irreducibilitasurol," *Mathematischer und Naturwissenschaftlicher Anzeiger der Ungarischen Akademie der Wissenschaften*, vol. 54 (1936), pp. 349-357.

[4] T. Nagell, "Über einige Irreduzibilitätskriterien," *Det Kongelige Norske Videnskabers Selskab Forhandlinger*, vol. 5 (1931), pp. 121-125.

[5] *Loc. cit.*, [1].

If $a_2 > 0$ *and if at least one of the following conditions*

$$\sqrt{a_2}(1 - P^{1/(n-1)}) \geqq \tfrac{1}{2} + \tfrac{1}{2}(2A - 1)^{\frac{1}{2}},$$

$$\sqrt{a_2}(1 - P^{1/(n-1)}) \geqq 1 + \{\tfrac{1}{2}(A + 1) - |a_1|\}^{\frac{1}{2}}$$

is satisfied, then the polynomial (1) *is irreducible in* P.

THEOREM II. *Assume that* $4a_2 - a_1^2 > 0$. *Set*

$$b = -a_1/2 + i(4a_2 - a_1^2)^{\frac{1}{2}}/2$$

and

$$Q = |a_3/b^3 + a_4/b^4 + \cdots + a_n/b^n|.$$

If

$$(4a_2 - a_1^2)/a_1^2 > Q^{2/n}/(1 - Q^{2/n})$$

and if at least one of the inequalities

$$\sqrt{a_2}(1 - Q^{1/n}) > \tfrac{1}{2} + \tfrac{1}{2}(2A - 1)^{\frac{1}{2}},$$

$$\sqrt{a_2}(1 - Q^{1/n}) > 1 + \{\tfrac{1}{2}(A + 1) - |a_1|\}^{\frac{1}{2}}$$

is satisfied, then the polynomial (1) *is irreducible in* P.

THEOREM III. *Set* $s = |a_1| + |a_3| + \cdots + |a_n|$. *The polynomial* (1) *is irreducible in* P *if* $a_2 \geqq 4^{2n-2}s^2$ *or if* $a_2 \geqq (7/2)^{2n-2}$ *for* $n \geqq 5$.

Perron's results were improved by Lipka [6] in two papers.

THEOREM IV. *The polynomial* (1) *is irreducible in* P *if* $a_2 > \min\{9s^2/2, 2(1+s)^2\}$.

THEOREM V. *The polynomial* (1) *is irreducible in* P *if* $a_1 = 0$ *and* $a_2 > 3s$.

In the following, these results will be improved furthermore.

THEOREM 1. *Let*

$$f(x) = x^n + a_1x^{n-1} + \cdots + a_n \qquad (a_n \neq 0) \tag{2}$$

be a polynomial with integral coefficients. Let m *be the minimum of the partial sums of the series*

$$0 + a_3 + a_4 + \cdots + a_n \tag{3}$$

---

[6] St. Lipka, "Über die Irreduzibilität von Polynomen," *Mathematische Annalen*, vol. 118 (1941-1943), pp. 235-245 and *loc. cit.*[3].

and m^* *the minimum of the partial sums of*

(4) $$0 - a_3 + a_4 - + \cdots + (-1)^n a_n.$$

Set

(5) $$t = |1 + a_4| + |a_1 + a_3| + |a_5| + |a_6| + \cdots + |a_n|.$$

If

(6) $$\begin{cases} a_2 > \max(t, a_1^2/4 + |m^*|) & \text{for } a_1 \geqq 0, \\ a_2 > \max(t, a_1^2/4 + |m|) & \text{for } a_1 \leqq 0, \end{cases}$$

then $f(x)$ *is irreducible in* P.

It is obvious that this result is sharper than Theorem III and IV since $|m| \leqq s - |a_1|$ and $|m^*| \leqq s - |a_1|$. For instance, let us consider the polynomials

$$f(x) = x^5 - 4x^4 + Kx^3 + 4x^2 - 2x - 2$$

with positive K. Theorem III is applicable if $K > 3{,}242{,}700$ and Theorem IV if $K > 338$. It is often tedious to see whether Theorem I and II can be used. Here they are certainly not applicable if $K < 100{,}000$ and $K < 100$, respectively.

It follows from Theorem 1 that $f(x)$ is irreducible if $K \geqq 5$ since $t = 3$ and $m = 0$. From the sharper Theorem 1a (see **3**) we obtain irreducibility for $K \geqq 4$, and $f(x)$ is reducible for $K = 3$ since $x = 1$ is one of the zeros. Therefore our results are the best possible for this special case.

A much more general case in which Theorem 1 gives the best possible result are the polynomials

$$f(x) = x^n - a_1 x^{n-1} + a_2 x^{n-2} - a_3 x^{n-3} - a_4 x^{n-4} \cdots - a_n$$

where all the a_ν are non-negative, $a_4 > 0$, and

$$a_1^2 \leqq 4(a_1 + a_3 + |a_4 - 1| + a_5 + \cdots + a_n).$$

Here we have $m = 0$, $t \geqq \frac{1}{4}a_1^2$, and $s = t + 1$. From Theorems III and IV we obtain irreducibility if

$$a_2 \geqq (7/2)^{2n-2}(t+1)^2 \text{ for } n \geqq 5 \text{ and } a_2 \geqq 2(t+2)^2,$$

respectively. These bounds for a_2 are very large if t is large. It follows from Theorem 1 that $f(x)$ is already irreducible if $a_2 > t$ since $t = \max(t, \frac{1}{4}a_1^2 + |m|)$, and for $a_2 = t$ the polynomial is reducible because $x = 1$ is a zero.

Theorem V will be improved as follows.

THEOREM 2. *The polynomial* (2) *is irreducible in* P *if* $a_1 = 0$ *and* $a_2 > t$.

Similar results hold if the coefficient a_{n-2} is large.

1. A lemma.

LEMMA. *Let*

$$g(x) = c_0x^k + c_1x^{k-1} + \cdots + c_k$$

be a polynomial with integral rational coefficients. Let m *be the smallest partial sum of the series*

$$c_{-1} + c_0 + c_1 + \cdots + c_k \tag{7}$$

where $c_{-1} = 0$. *Then we have for* $x > 1$

$$g(x) \geqq mx^k.$$

Proof. It follows from (7) that

$$m \leqq 0. \tag{8}$$

Since the coefficients c_κ are rational integers, we may write each positive term $c_\kappa x^\kappa$ as the sum of c_κ terms x^κ and each negative term $c_\kappa x^\kappa$ as the sum of $| c_\kappa |$ terms $-x^\kappa$. Assume that r is the greatest integer such that

$$c_{-1} + c_0 + c_1 + \cdots + c_r = m. \tag{9}$$

We consider the terms $\pm x^{k-\rho}$ with $0 \leqq \rho \leqq r$. For $r = -1$ such terms do not exist. Since m was the minimum of the partial sums of (7), it follows from (8) and (9) that to each positive term $x^{k-\alpha}$ with $0 \leqq \alpha < r$ we may assign a negative term $-x^{k-\beta}$ with $\alpha < \beta \leqq r$. Then

$$x^{k-\alpha} - x^{k-\beta} > 0 \qquad \text{for } x > 1 \tag{10}$$

and the sum S of all these pairs of corresponding terms is certainly not negative. By (8) and (9), we have exactly $| m |$ such negative terms more than positive terms. Therefore exactly $| m |$ negative terms do not belong to one of the pairs (10). Hence by (8)

$$\sum_{\rho=0}^{r} c_\rho x^{k-\rho} \geqq S - | m | x^k \geqq mx^k. \tag{11}$$

This proves the Lemma if $r = k$.

Assume now that $r < k$. We consider the terms $\pm x^{k-\rho}$ with $r < \rho \leqq k$.

It follows from (9) that to each negative term $-x^{k-\delta}$ with $r < \delta \leqq k$ we may assign a positive $x^{k-\gamma}$ with $r < \gamma < \delta$. Hence

$$(12) \qquad x^{k-\gamma} - x^{k-\delta} > 0 \text{ for } x > 1$$

and the sum extended over all the pairs (12) is positive. Hence

$$(13) \qquad \sum_{\rho=r+1}^{k} c_\rho x^{k-\rho} > 0 \text{ for } x > 1 \text{ and } r < k.$$

Adding (11) and (13) we obtain our Lemma for $r < k$.

Corollary. *If $g(x)$ does not vanish identically, then*

$$g(x) > mx^{k+1} \textit{ for } x > 1.$$

Proof. The statement follows from (11) and (13) if $r < k$. Assume now that $r = k$. If at least one of the $c_\rho > 0$ where $\rho \leqq r$, then $S > 0$. Otherwise $m < 0$, hence for $x > 1$, $S - |m| x^k > mk^{k+1}$, and the proof follows from (11).

2. Proof of Theorem 1. We shall use the following theorem of Berwald.[7]

Let

$$F(x) = c_0 x^n + c_1 x^{n-1} + \cdots + c_n$$

be a polynomial with real coefficients, p an integer such that $0 \leqq p \leqq n$, and $l = \max(p, n-p)$. If

$$|c_{n-p}| > \sum_{\lambda=1}^{l} |c_{n-p+\lambda} + c_{n-p-\lambda}|$$

where $c_\sigma = 0$ for $\sigma < 0$ and $\sigma > n$, then $F(x)$ has exactly p zeros inside the unit circle and no root on its contour.

Proof of Theorem 1. Since by (5) and (6)

$$a_2 > |a_1 + a_3| + |1 + a_4| + |a_5| + \cdots + |a_n| = t,$$

it follows from the theorem of Berwald for $p = n-2$ that $f(x)$ has exactly $n-2$ roots inside the unit circle and the two remaining roots x_1 and x_2 lie in the exterior.

If $f(x)$ is reducible in P, then x_1 and x_2 must be real numbers. Other-

[7] *Loc. cit.*[2].

wise they must be conjugate complex numbers and therefore roots of the same factor of $f(x)$. All the roots of the other factor lie in the interior of the unit circle and are different from 0 since $a_n \neq 0$. Hence their product cannot be a rational integer.

Therefore we only have to prove that x_1 and x_2 are complex numbers. It follows from (5) that

$$(15) \qquad t \geqq | a_4 | - 1 + | a_3 | - | a_1 | + | a_5 | + | a_6 | + \cdots + | a_n |.$$

By (3), either $m = 0$, or there exists an integer $\nu \leqq n$ such that

$$m = a_3 + a_4 + \cdots + a_\nu,$$

hence

$$(16) \quad | m | \leqq | a_3 | + | a_4 | + \cdots + | a_\nu | \leqq | a_3 | + | a_4 | + \cdots + | a_n |.$$

It follows from (6), (15), and (16) that

$$(17) \qquad 1 + | a_1 | + a_2 - | m | > 1 + | a_1 | + t - | m | \geqq 0.$$

In exactly the same way we obtain

$$(18) \qquad 1 + | a_1 | + a_2 - | m^* | > 1 + | a_1 | + t - | m^* | \geqq 0.$$

We set

$$(19) \qquad \begin{cases} g(x) = a_3x^{n-3} + a_4x^{n-4} + \cdots + a_n, \\ h(x) = -a_3x^{n-3} + a_4x^{n-4} - + \cdots + (-1)^n a_n. \end{cases}$$

It follows from the Lemma, by (3) and (4), that for $x > 1$

$$(20) \qquad a_2x^{n-2} + g(x) \geqq a_2x^{n-2} + mx^{n-3} \geqq (a_2 - | m |)x^{n-2},$$

$$(21) \qquad a_2x^{n-2} + h(x) \geqq a_2x^{n-2} + m^*x^{n-3} \geqq (a_2 - | m^* |)x^{n-2},$$

$$(22) \qquad -a_2x^{n-2} - h(x) \leqq (-a_2 + | m^* |)x^{n-2}.$$

We want to prove now that $f(x)$ has no positive root $\xi > 1$. We have to distinguish between two cases.

1) $$a_1 \geqq 0.$$

By (2), (19), and (20) we have

$$(23) \quad \begin{cases} f(\xi) = \xi^n + a_1\xi^{n-1} + a_2\xi^{n-2} + g(\xi) \geqq \xi^n + a_1\xi^{n-1} + (a_2 - | m |)\xi^{n-2} \\ \qquad = \xi^n + | a_1 | \xi^{n-1} + (a_2 - | m |)\xi^{n-2} > (1 + | a_1 | + a_2 - | m |)\xi^{n-2}, \end{cases}$$

hence by (17)

$$f(\xi) > 0 \text{ for } \xi > 1.$$

2) $$a_1 < 0.$$

Here it follows from (2), (19), and (20) that

$$(24)\quad \begin{cases} f(\xi) = \xi^n + a_1\xi^{n-1} + a_2\xi^{n-2} + g(\xi) \geqq \xi^n + a_1\xi^{n-1} + (a_2 - |m|)\xi^{n-2} \\ \qquad = \xi^{n-2}(\xi^2 + a_1\xi + a_2 - |m|). \end{cases}$$

The discriminant $a_1^2 - 4(a_2 - |m|)$ of the quadratic polynomial in parentheses is negative by (6). Hence the quadratic polynomial is positive definite and $f(\xi)$ is positive for $\xi > 1$.

Therefore in both cases $f(x)$ cannot have a positive root $\xi > 1$.

Now we assume that $\xi < -1$. We set $\xi = -\eta$. Here we have to distinguish between 4 cases.

1) n is even and $a_1 \geqq 0$.

We have by (19) and (21) for $\eta > 1$

$$(25)\quad \begin{cases} f(\xi) = f(-\eta) = \eta^n - a_1\eta^{n-1} + a_2\eta^{n-2} + h(\eta) \\ \qquad \geqq \eta^n - a_1\eta^{n-1} + (a_2 - |m^*|)\eta^{n-2} = \eta^{n-2}(\eta^2 - a_1\eta + a_2 - |m^*|); \end{cases}$$

since the quadratic polynomial in parentheses is positive definite by (6). Hence $f(\xi)$ is positive for $\xi < -1$.

2) n is odd and $a_1 \geqq 0$.

It follows from (19) and (22) that

$$(26)\quad \begin{cases} f(\xi) = f(-\eta) = -\eta^n + a_1\eta^{n-1} - a_2\eta^{n-2} - h(\eta) \\ \qquad \leqq -\eta^n + a_1\eta^{n-1} + (-a_2 + |m^*|)\eta^{n-2} = \\ \qquad -\eta^{n-2}(\eta^2 - a_1\eta + a_2 - |m^*|) < 0 \end{cases}$$

since the quadratic polynomial in parentheses is again positive definite. In this case $f(\xi)$ is negative for $\xi < -1$.

3) n is even and $a_1 < 0$.

We have by (25)

$$f(\xi) = f(-\eta) \geqq \eta^n - a_1\eta^{n-1} + (a_2 - |m^*|)\eta^{n-2}.$$

Since $a_1 < 0$, it follows that

$$(27) \qquad \begin{aligned} f(\xi) &\geqq \eta^n + |a_1|\eta^{n-1} + (a_2 - |m^*|)\eta^{n-2} \\ &> (1 + |a_1| + a_2 - |m^*|)\eta^{n-2}, \end{aligned}$$

hence by (18)

$$f(\xi) > 0 \text{ for } \xi < -1.$$

4) n is odd and $a_1 < 0$.

Here we have by (26)

$$(28) \qquad \begin{aligned} f(\xi) = f(-\eta) &\leqq -\eta^n + a_1\eta^{n-1} - (a_2 - |m^*|)\eta^{n-2} \\ &= -\eta^n - |a_1|\eta^{n-1} + (a_2 - |m^*|)\eta^{n-2} \\ &< -\eta^{n-2}(1 + |a_1| + a_2 - |m^*|) \end{aligned}$$

and by (18)

$$f(\xi) < 0 \text{ for } \xi < -1.$$

Therefore $f(\xi) \neq 0$ for $\xi < -1$ in each case. This proves our theorem.

3. Corollaries. If instead of (6) we only assume that

$$a_2 = t \geqq a_1^2/4 + m^* \text{ for } a_1 \geqq 0,$$

then Berwald's theorem does not hold anymore. The polynomial $f(x)$ may have roots on the boundary of the unit circle and may be reducible in P. Examples of such polynomials were given already in the introduction.

If instead of (6) we assume that

$$a_2 = a_1^2/4 + |m^*| \text{ for } a_1 \geqq 0$$

or

$$a_2 = a_1^2/4 + |m| \text{ for } a_1 \leqq 0,$$

then again $f(x)$ may become reducible in P. This can be seen by the examples

$$x^2 + 2bx + b^2 \text{ with } |b| > 2.$$

Here we have $m = m^* = 0$ and $t = 1 + |2b| < b^2 = a_1^2/4$, hence

$$a_2 = b^2 = \max(1 + |2b|, b^2) = \max(t, a_1^2/4).$$

But the polynomial is irreducible in P if $n > 2$. We have

THEOREM 1a. *If the polynomial* (2) *satisfies the condition*

$$a_2 \geqq a_1^2/4 + |m^*| > t \textit{ for } a_1 \geqq 0$$

or

$$a_2 \geqq a_1^2/4 + |m| > t \textit{ for } a_1 \leqq 0$$

where t, m, and m^* are defined by (3), (4), *and* (5), *and if $n > 2$, then $f(x)$ is irreducible in* P.

Proof. It is easy to see that the inequalities (23), (27), and (28) remain correct here. Therefore in these cases

(29) $$f(\xi) \neq 0 \text{ for } \xi > 1 \text{ and for } \xi < -1.$$

The discriminants of the quadratic polynomials of (24), (25), and (26) vanish now. But since $n > 2$, the polynomials $g(x)$ and $h(x)$ are not identically zero, and we obtain

$$a_2x^{n-2} + g(x) > a_2x^{n-2} + mx^{n-2} = (a_2 - |m|)x^{n-2},$$
$$a_2x^{n-2} + h(x) > a_2x^{n-2} + m^*x^{n-2} = (a_2 - |m^*|)x^{n-2}$$

instead of (20) and (21) by the Corollary of the Lemma; therefore

$$f(\xi) > \xi^{n-2}(\xi^2 + a_1\xi + a_2 - |m|) \geqq 0,$$
$$f(\xi) > \eta^{n-2}(\eta^2 - a_1\eta + a_2 - |m^*|) \geqq 0,$$
$$f(\xi) < -\eta^{n-2}(\eta^2 - a_1\eta + a_2 - |m^*|) \leqq 0$$

instead of (24), (25), and (26), respectively. Hence (29) remains correct.

We are now able to prove Theorem 2.

Proof. For $n = 2$, the theorem is trivial since a polynomial of form $x^2 + a_2$ with $a_2 > 0$ is irreducible in P. Let us assume now that $n > 2$. It follows from Theorems 1 and 1a that $f(x)$ is irreducible if $a_2 > t$ and $a_2 \geqq |m|$.

On the other hand, we have by (17)

(30) $$|m| \leqq 1 + t$$

and by assumption

(31) $$a_2 > t,$$

hence by (30)

(32) $$a_2 \geqq 1 + t \geqq |m|$$

since a_2 and t are integers. Now Theorem 2 follows by (31) and (32).

THEOREM 3. *Let*

$$f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-2}x^2 + a_{n-1}x + a_n \qquad (a_0a_n \neq 0) \tag{33}$$

be a polynomial with integral rational coefficients. Denote the minima of the partial sums of

$$0 + a_{n-3}a_n^2 + a_{n-4}a_n^3 + \cdots + a_1a_n^{n-2} + a_0a_n^{n-1}$$

and

$$0 - a_{n-3}a_n^2 + a_{n-4}a_n^3 - + \cdots + (-1)^{n-1}a_1a_n^{n-2} + (-1)^n a_0a_n^{n-1}$$

by μ *and* μ^*, *respectively. Set*

$$\tau = |1 + a_{n-4}a_n^3| + |a_{n-1} + a_{n-3}a_n^2| + \sum_{\nu=5}^{n} |a_{n-\nu}a_n^{\nu-1}|.$$

If

$$a_{n-2}a_n > \max(\tau, a_{n-1}^2/4 + |\mu^*|) \quad \textit{for } a_{n-1} > 0,$$

$$a_{n-2}a_n > \max(\tau, a_{n-1}^2/4 + |\mu|) \quad \textit{for } a_{n-1} < 0,$$

then $f(x)$ *is irreducible in* P.

Proof. We set

$$F(x) = x^n f(x^{-1}) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0,$$

$$y = a_nx,$$

and

$$G(y) = a_n^{n-1}F(x) = y^n + a_{n-1}y^{n-1} + a_{n-2}a_ny^{n-2} + \cdots + a_1a_n^{n-2}y + a_0a_n^{n-1}.$$

It follows from Theorem 1 that $G(y)$ is irreducible in P. Hence $F(x)$ and $f(x)$ are also irreducible in P.

In the same way we obtain from Theorem 2

THEOREM 4. *The polynomial* (33) *is irreducible in* P *if* $a_{n-1} = 0$ *and* $a_{n-2}a_n > \tau$.

UNIVERSITY OF NORTH CAROLINA.

ELASTICITY BEYOND THE ELASTIC LIMIT.*

By M. Reiner.

---

1. Theories of elasticity have so far presupposed the existence of what Love (Art. 76) called a "state of ease" of "perfect elasticity" in which "a body can be strained without taking any set"; that state ranging between an "initial," "unstressed" and "unstrained" state (Art. 64) on one hand and the "elastic limit" on the other. Recent technological progress has gradually reduced, absolutely and, still more, relatively, the field in which this assumption holds good. Not only has increased accuracy of measurements of permanent sets lowered the elastic limit until in many cases as, for instance, annealed copper, it has nearly disappeared. More important, in materials which *do* show a definite elastic limit as, for instance, mild steel, deformations in most practical applications go *beyond* that limit. In addition, one has to consider elastic materials such as bitumen or cement-stone showing creep: their elastic potential gradually disappears through relaxation. Finally, there are such materials as rubber which can be caused to undergo very large deformations, a certain part of which will always be non-recoverable. It therefore becomes necessary to consider elasticity beyond the elastic limit. If we define elasticity with Love as "the property of recovery of an *original* size and shape," there would in all these cases be no question of elasticity because the *original* size and shape is not recovered. However, *some* of the deformation is *always* recovered: but *which* part of it is recoverable, becomes apparent only when all external forces, gravity included, *have been removed.* We may denote as the *ground-position* that position of the body which is then reached. To every deformation there corresponds a ground-position of its own, which generally will not be the *initial* position from which the deformation started. Let us denote by *deformation* a change of size and of shape in general, whether recoverable or not, and by *strain* that part of it which is recovered when all external forces have been removed. Generally, the strain will differ from the deformation not only in magnitude, but also in the orientation of the principal axes.

The ground-position is accordingly an unstrained and unstressed state, but it is not an *undeformed* state. A general theory of elasticity, then, has to relate the strain as now defined to (i) the stress produced by it and (ii) the

---

* Received October 12, 1947.

external forces necessary to equilibrate the stresses in the body in accordance with d'Alembert's principle; while the classical theory of the "state of ease" refers to the special case when the strain is identical with the deformation. The considerations of the present paper are, however, also applicable in the latter case.

2. The classical theory was brought to completion by Murnaghan when considering finite strain. He derived the relation between the stress tensor T^{rs} and the strain tensor ϵ_{rs} from a formula connecting the elastic potential ϕ with the stress tensor.[1] That formula itself was derived by considering the virtual work of the stresses across a closed boundary of a portion of the material. This method is inapplicable in our case. If we write the fundamental law of thermodynamics for isothermal processes in the form of the Gibbs-Helmholtz equation

$$\delta w = \rho\delta\phi + \rho\delta\psi \tag{2.1}$$

where w is the strainwork per unit volume, ϕ the intrinsic free energy-density and ψ the bound energy-density (compare Weissenberg, 1931), not only will ψ in our general case not vanish, but what is more remarkable, as Taylor and Quinney have shown in a metal which is subjected to cold working, part of the free energy is "latent" and not recoverable mechanically. We therefore must apply that other method used in the classical theory for the derivation of the stress-strain relation (e.g. by Stokes) which is a generalization of Hooke's law, writing

$$T_s{}^r = f(\epsilon_s{}^r) \tag{2.2}$$

and developing the function f by means of tensor analysis, as was done by Reiner in the analogous case of viscosity. The equation will then express a law of elasticity if ϵ indicates the strain defined above as the recovered part of the deformation and if the relation connecting $T_s{}^r$ and $\epsilon_s{}^r$ is unequivocal. From the last condition there follows, that we can also write

$$\epsilon_s{}^r = \mathfrak{f}(T_s{}^r). \tag{2.3}$$

In the experimental determination of the relation one would have in principle to proceed as follows: Subject a material to external forces and let it undergo a process of deformation of a certain type,[2] arrest the deformation and record

[1] We shall use in the present paper wherever possible Murnaghan's notation.

[2] For "type" compare Love (Art. 73).

the magnitude of stress;[3] mark a sphere of unit radius in the material around some selected point; remove all external forces: this will induce relative displacements in the material changing the sphere into an ellipsoid called the reciprocal strain ellipsoid; wait until this movement dies out; measure the axes of the ellipsoid: they will provide a measure of strain; repeat this experiment reaching different magnitudes of the same type of deformation in a gradually increasing or decreasing order; record the strain, as determined, against the stress: provided the relaxation of the stress is negligible, the result is an empirical relation for (2.3) depending upon the type of deformation.[4] For instance, in the usual tensile test for metals in the work-hardening range, when the volume of an element of the material can be assumed as constant and the deformation has axial symmetry, only one axis of the ellipsoid need be measured *viz.*, either along or across the test piece and the empirical formula relates the axial traction p_{zz} to the axial strain ϵ_{zz}.

3. In the first stage of our investigation we need not fix the *measure* of strain. Denoting the three axes of the above mentioned ellipsoid by l_i (i running over 1, 2, 3) or, what is sometimes more convenient, the axes of the strain ellipsoid by $\lambda_i = 1/l_i$ "one may use (as Weissenberg (1946) pointed out) any function of the elongation ratios (λ_i) in the direction of the main axes choosing the function to suit the particular field of investigation." One would, naturally, require that all these functions are reduced for infinitesimal strain to the Cauchy measure $\lambda_i - 1$. This is the case with the Kirchoff-measure which is based upon $\frac{1}{2}[(\lambda_i)^2 - 1]$ and the Murnaghan-measure based upon $\frac{1}{2}[1 - (l_i)^2]$; it is also so with the measure $ln\ (\lambda_i) = -ln(l_i)$ originally proposed by Roentgen for rubber and, since systematically introduced by Hencky, now widely in use. We may also mention the measure $(\lambda_i - l_i)$ proposed by Wall. All these measures comply also with a second requirement, *viz.* that the strain vanishes for $\lambda_i = 1 = l_i$. It is clear that a linear stress-strain relation in one measure will be non-linear in every other and the desire for linearity is often one of the motives behind the introduction of one or the other of the measures mentioned, our enumeration being far from complete.

4. Starting from (2.2) or (2.3), we follow the reasoning applied by Reiner, as has already been mentioned, in the analogous case of viscous

[3] It is necessary first to arrest the deformation as, generally, part of the stress will be due to viscous resistance, depending upon the velocity of deformation.

[4] "If . . . the stress-strain relations can be found experimentally, the strain-energy function can be calculated" (Sokolnikoff, p. 89).

resistance. We note that on the left side there stands a mixed tensor of rank two. Then in a development of either function f or $\mathfrak{f}$ all terms on the right side must also be mixed tensors of rank two. The right side can therefore consist only of sums of mixed tensors of rank two multiplied by scalars and of inner products of such tensors which again are reduced to tensors of rank two. The general term of a development of the function f will therefore be of the form $\epsilon_\alpha{}^r\epsilon_\beta{}^\alpha\epsilon_\gamma{}^\beta\cdots\epsilon_s{}^\lambda\cdot f(I)$, where $f(I)$ is a function of the three invariants: and we can therefore write

$$(4.1)\qquad T_s{}^r = f_0\delta_s{}^r + f_1\epsilon_s{}^r + f_2\epsilon_\alpha{}^r\epsilon_s{}^\alpha + f_3\epsilon_\alpha{}^r\epsilon_\beta{}^\alpha\epsilon_s{}^\beta + \cdots.$$

This would mean an infinite number of such terms. However, in view of the Cayley-Hamilton equation of matrix theory, the following relation holds good [5]

$$(4.2)\qquad \epsilon_\alpha{}^r\epsilon_\beta{}^\alpha\epsilon_s{}^\beta = \delta_s{}^r III - \epsilon_s{}^r II + \epsilon_\alpha{}^r\epsilon_s{}^\alpha I$$

where I, II and III are the first, second and third invariants respectively.

Therefore

$$(4.3)\qquad \begin{aligned}\epsilon_\alpha{}^r\epsilon_\beta{}^\alpha\epsilon_\gamma{}^\beta\epsilon_s{}^\gamma &= \delta_\alpha{}^r\epsilon_s{}^\alpha III - \epsilon_\alpha{}^r\epsilon_s{}^\alpha II + \epsilon_\alpha{}^r\epsilon_\beta{}^\alpha\epsilon_s{}^\beta I\\ &= \delta_s{}^r I\cdot III + \epsilon_s{}^r(III - I\cdot II) + \epsilon_\alpha{}^r\epsilon_s{}^\alpha(I^2 - II)\end{aligned}$$

and similarly with respect to higher terms. This enables us to write

$$(4.4)\qquad T_s{}^r = F_0\delta_s{}^r + F_1\epsilon_s{}^r + F_2\epsilon_\alpha{}^r\epsilon_s{}^\alpha$$

and analogously

$$(4.5)\qquad \epsilon_s{}^r = \mathfrak{F}_0\delta_s{}^r + \mathfrak{F}_1T_s{}^r + \mathfrak{F}_2T_\alpha{}^rT_s{}^\alpha$$

where the F are functions of the three invariants I_ϵ, II_ϵ and III_ϵ of the strain-, and the $\mathfrak{F}$ functions of the three invariants I_T, II_T and III_T of the stress-tensor. Prager has recently derived equations built up in a manner similar to (4.4) and (4.5), but subject to specializations due to certain simplifying assumptions. Our equations are general and express nothing more than that both stress and strain are tensors of rank two, the principal axes of which coincide; and that the functions F and $\mathfrak{F}$ are scalars. We may call a material in such a state *isotropic*.

However, we also require that in the ground-position, when the stress is removed, the strain should also vanish, and *vice versa*. Therefore

---

[5] These developments are entirely analogous to those of Reiner for the viscous liquids, but it was thought desirable to make the present paper self-contained.

(4. 6) $$F_0 = F_{01} I_\epsilon + F_{02} II_\epsilon + F_{03} III_\epsilon$$

(4. 7) $$\mathfrak{F}_0 = \mathfrak{F}_{01} I_T + \mathfrak{F}_{02} II_T + \mathfrak{F}_{03} III_T$$

where the new F and $\mathfrak{F}$ are again, in general, functions of all three invariants I, II and III.

The functions F are *moduli* of elasticity, the functions $\mathfrak{F}$ *coefficients* of elasticity; the latter, generally, *not* the reciprocals of the former. There are therefore, generally, five of each kind, each one possessing ∞^3 values in accordance with the values which the invariants may have in every particular case. In the expressions for F and $\mathfrak{F}$ as functions of the invariants, there will appear a number of parameters, which are the elastic "constants" of the material. The F and $\mathfrak{F}$ may, of course, themselves be constants; in special cases some of them may vanish, in other cases they may not be independent; and this would reduce their number from five to less.

The F and $\mathfrak{F}$ can be given physical interpretations only when a definite measure of strain is assumed and we shall examine what consequences the adoption of any such measure may have.

5. Before dealing with the problem in a general way, it will be useful to examine the special case of simple shear dealt with by Love in Art. 37. This is given kinematically by the equations

(5. 1) $$x_1 = x + sy; \qquad y_1 = y; \qquad z_1 = z.$$

Putting

(5. 2) $$s = 2 \tan \alpha,$$

Love calculates

(5. 3) $$\lambda_1 = \frac{1 - \sin \alpha}{\cos \alpha}; \qquad \lambda_2 = \frac{1 + \sin \alpha}{\cos \alpha}; \qquad \lambda_3 = 1$$

and he proves that the directions of the principal axes of strain are the bisectors of the angle $(\pi/2) + \alpha$ with the x-axis, and the angle through which the principal axes are turned is the angle α. The stress caused by the strain will have the principal components T_1, T_2, T_3 which from (4. 4) and (4. 6) are

(5. 4) $$T_i = F_{01} I + F_{02} II + F_{03} III + F_1 \epsilon_i + F_2 \epsilon_i^2.$$

The components of stress with respect to the system x, y, z will be from Love's equations, Art. 49:

$$(5.5)\quad \begin{aligned} T_{xx} &= \tfrac{1}{2}(T_1+T_2) - \tfrac{1}{2}(T_1-T_2)\sin\alpha \\ T_{yy} &= \tfrac{1}{2}(T_1+T_2) + \tfrac{1}{2}(T_1-T_2)\sin\alpha \\ T_{zz} &= T_3 \\ T_{xy} &= -\tfrac{1}{2}(T_1-T_2)\cos\alpha; \quad T_{yz} = T_{zx} = 0. \end{aligned}$$

Introducing the expressions for the principal stresses from (5. 4) into (5. 5) gives

$$(5.6)\quad \begin{aligned} T_{xx} &= F_{01}I + F_{02}II + F_{03}III + \tfrac{1}{2}\{F_1[(\epsilon_1+\epsilon_2) - (\epsilon_1-\epsilon_2)\sin\alpha] \\ &\qquad + F_2[(\epsilon_1^2+\epsilon_2^2) - (\epsilon_1^2-\epsilon_2^2)\sin\alpha]\} \\ T_{yy} &= F_{01}I + F_{02}II + F_{03}III + \tfrac{1}{2}\{F_1[(\epsilon_1+\epsilon_2) + (\epsilon_1-\epsilon_2)\sin\alpha] \\ &\qquad + F_2[(\epsilon_1^2+\epsilon_2^2) + (\epsilon_1^2-\epsilon_2^2)\sin\alpha]\} \\ T_{zz} &= F_{01}I + F_{02}II + F_{03}III \\ T_{xy} &= -\tfrac{1}{2}[F_1(\epsilon_1-\epsilon_2) + F_2(\epsilon_1^2-\epsilon_2^2)]\cos\alpha. \end{aligned}$$

We now assume definite measures of strain. If l_0 is a length extended in simple elongation by Δl to l, the measure of the extension may relate Δl to either l_0 or l, or it may relate an element of elongation dl to the instantaneous length l. These three possibilities correspond to the Kirchhoff-measure.

$$(5.7)\qquad \epsilon_i^K = \tfrac{1}{2}(\lambda_i^2 - 1)$$

the Murnaghan-measure

$$(5.8)\qquad \epsilon_i^M = \tfrac{1}{2}(1 - l_i^2) = \tfrac{1}{2}(1 - 1/\lambda_i^2)$$

and to the logarithmic or Hencky-measure

$$(5.9)\qquad \epsilon_i^H = ln\lambda_i = -\,lnl_i.$$

Introducing the expressions λ_i from (5. 3), we find the principal strain-components, the strain-invariants and the stress-components in the x, y and z directions as entered in the following Table:

ϵ_i	$\frac{1}{2}(\lambda_i^2 - 1)$	$ln\lambda_i$	$\frac{1}{2}(1 - 1/\lambda_i^2)$
ϵ_1	$-\tan\alpha \frac{1-\sin\alpha}{\cos\alpha}$	$ln \frac{1-\sin\alpha}{\cos\alpha}$	$-\tan\alpha \frac{\cos\alpha}{1-\sin\alpha}$
ϵ_2	$\tan\alpha \frac{\cos\alpha}{1-\sin\alpha}$	$ln \frac{\cos\alpha}{1-\sin\alpha}$	$\tan\alpha \frac{1-\sin\alpha}{\cos\alpha}$
ϵ_3	0	0	0
I_ϵ	$2(\tan\alpha)^2$	0	$-2(\tan\alpha)^2$
II_ϵ	$-(\tan\alpha)^2$	$-(ln \frac{\cos\alpha}{1-\sin\alpha})^2$	$-(\tan\alpha)^2$
III_ϵ	0	0	0
T_{xx}	$(\tan\alpha)^2[2F_{01} - F_{02} + 2F_1 + F_2(1 + 4\tan^2\alpha)]$	$ln \frac{\cos\alpha}{1-\sin\alpha}[(F_2 - F_{02}) ln \frac{\cos\alpha}{1-\sin\alpha} + F_1 \sin\alpha]$	$-(\tan\alpha)^2(2F_{01} + F_{02} - F_2)$
T_{yy}	$(\tan\alpha)^2(2F_{01} - F_{02} + F_2)$	$ln \frac{\cos\alpha}{1-\sin\alpha}[(F_2 - F_{02}) ln \frac{\cos\alpha}{1-\sin\alpha} - F_1 \sin\alpha]$	$-(\tan\alpha)^2[2F_{01} + F_{02} + 2F_1 - F_2(1 + 4\tan^2\alpha)]$
T_{zz}	$(\tan\alpha)^2(2F_{01} - F_{02})$	$-(ln \frac{\cos\alpha}{1-\sin\alpha})^2 F_{02}$	$-(\tan\alpha)^2(2F_{01} + F_{02})$
T_{xy}	$\tan\alpha(F_1 + 2F_2\tan^2\alpha)$	$ln \frac{\cos\alpha}{1-\sin\alpha} F_1$	$\tan\alpha(F_1 - 2F_2\tan^2\alpha)$

In infinitesimal strain we may neglect $(\tan\alpha)^2$ and introducing $\tan\alpha = \alpha = s/2$ all three measures give the same $\epsilon_2 = -\epsilon_1 = s/2$, $T_{xy} = F_1 s/2$, while normal tractions T_{xx}, T_{yy} and T_{zz} vanish. In this case simple shear is accompanied by a shearing stress only. In finite strain such a simple relation is not possible, whatever the values of the elastic moduli. We may in the Kirchhoff measure make $F_{02} = 2F_{01}$ and T_{zz} vanishes. We may in addition make $F_2 = 0$ and T_{yy} will vanish. But there *must* remain a tension in the direction x which is $T_{xx} = 2F_1(\tan\alpha)^2$ and we cannot put $F_1 = 0$ because then T_{xy} also would vanish. Alternately, we may put $F_2 = -2F_1/[1 + 4(\tan\alpha)^2]$ which would make T_{xx} vanish, but leave a pressure $T_{yy} = -2F_1(\tan\alpha)^2/[1 + 4(\tan\alpha)^2]$. Conditions are similar in other measures. In an isotropic material finite simple shear is accompanied by either a tension in the direction of the displacement or compression in the direction of its gradient or both. Weissenberg (1947) has demonstrated the existence of such stresses in elastic liquids in a series of striking experiments.

6. The present theory is distinguished from the usual theory of elasticity of finite strains mainly by the appearance of the modulus F_2. In the usual theory, Equation (5.4) would be

$$T_i = F_0 + F_1\epsilon_i \tag{6.1}$$

which constitutes three equations with two unknowns, *viz.*, the moduli F_0 and F_1. In order that these equations should be consistent, certain relations between T_i and ϵ_i must be satisfied. The matrix of the coefficients is of rank two. The augmented matrix

$$K \equiv \begin{Vmatrix} 1 & \epsilon_1 & T_1 \\ 1 & \epsilon_2 & T_2 \\ 1 & \epsilon_3 & T_3 \end{Vmatrix} \tag{6.2}$$

must therefore also be of the rank two. This requires the determinant

$$\begin{vmatrix} 1 & \epsilon_1 & T_1 \\ 1 & \epsilon_2 & T_2 \\ 1 & \epsilon_3 & T_3 \end{vmatrix} = 0 \tag{6.3}$$

or

$$\frac{T_1 - T_2}{\epsilon_1 - \epsilon_2} = \frac{T_2 - T_3}{\epsilon_2 - \epsilon_3} = \frac{T_3 - T_1}{\epsilon_3 - \epsilon_1} = F(I_\epsilon, II_\epsilon, III_\epsilon, I_T, II_T, III_T). \tag{6.4}$$

Equation (6.4) has been proposed by Weissenberg (1947) as a law of elasticity. As has, however, been shown here, it is not general enough and is not

independent of the *measure* of strain. For instance, should experiments show that simple shear is accompanied by a tension in the direction of displacement, the Murnaghan measure could not be used. On the other hand, should experiments show that it is accompanied by compression in the direction of the gradient of the displacement, the Kirchhoff measure could not be used. In the form of Equation (5.4) the law of elasticity is independent of the measure and does not prejudice the outcome of experiment.

7. Considering that, by including the modulus F_2 (or the coefficient $\mathcal{F}_2$), we are independent of the measure of strain, we may for our further investigation assume any measure. We shall select the Hencky-measure for two reasons:

(i) Because of $\epsilon_i = ln\lambda_i$, $\dot{\epsilon}_i = \dot{\lambda}_i/\lambda_i$. Denoting by e_i the principal "velocity-extension" of hydrodynamics, we accordingly get $e_i = \dot{\epsilon}_i$, provided the principal axes do not rotate. Therefore in *pure* strain, in the Hencky-measure, to use Murnaghan's words "the variation of the strain tensor (is equal) to the space derivative of the virtual displacement vector." This is of advantage, especially if we consider that it may be possible in many cases to arrange "the removal of the external forces" (compare **2** and **3** above) in such a way that the axes do not rotate and the strain is acordingly pure.

(ii) Secondly, from

$$V/V_0 = \lambda_1 \cdot \lambda_2 \cdot \lambda_3, \tag{7.1}$$

there follows

$$\epsilon_v = ln(V/V_0) = ln\lambda_1 + ln\lambda_2 + ln\lambda_3 = \epsilon_1 + \epsilon_2 + \epsilon_3 = I_\epsilon. \tag{7.2}$$

Therefore, in the Hencky measure, and only in that measure, the cubical dilation is equal to the first invariant of the strain tensor.[6] Accordingly, only in this measure has the resolution of the tensor in an isotropic and a deviatoric component physical significance.

8. By carying out the resolutions

$$T_s{}^r = T\delta_s{}^r + T'_s{}^r; \qquad \epsilon_s{}^r = \epsilon\delta_s{}^r + \epsilon'_s{}^r \tag{8.1}$$

where

$$T_a{}^a = 3T, \quad \epsilon_a{}^a = 3; \qquad T'_a{}^a = \epsilon'_a{}^a = 0 \tag{8.2}$$

we get from (4.4) and (4.6)

[6] This is, of course, also the case in infinitesimal strain.

$$(8.3)\qquad \begin{aligned} T &= F_{01} + F_{02} + II'_\epsilon + F_{03}III'_\epsilon \\ T'_s{}^r &= F_1\epsilon'_s{}^r + F_2(\epsilon'_a{}^r\epsilon'_s{}^a + 2II'_\epsilon/3\cdot\delta_s{}^r) \end{aligned}$$

and from (4. 5) and (4. 7)

$$(8.4)\qquad \begin{aligned} \epsilon &= \mathfrak{F}_{01}T + \mathfrak{F}_{02}II'_T + \mathfrak{F}_{03}III'_T \\ \epsilon'_s{}^r &= \mathfrak{F}_1 T'_s{}^r + \mathfrak{F}_2(T'_a{}^r T'_s{}^a + 2II'_T/3\cdot\delta_s{}^r) \end{aligned}$$

where the accents indicate the deviator and the F and $\mathfrak{F}$ are now functions of the invariants of the deviator, different from the functions F and $\mathfrak{F}$ appearing in (4. 4) to (4. 7).

If we introduce $\epsilon'_s{}^r$ from (8. 4) into (8. 3), considering that $\delta_s{}^r$, $T'_s{}^r$ and $T'_a{}^r T'_s{}^a$ stand for the zero, first and second powers in the stress components, we find

$$(8.5)^{7}\qquad F_1 = \frac{\mathfrak{F}_1{}^2 + II'_T/3\cdot\mathfrak{F}_2{}^2}{\begin{vmatrix} \mathfrak{F}_1[\mathfrak{F}_2 III'_T - 2II'_T/3\cdot\mathfrak{F}_1]\mathfrak{F}_2 \\ \mathfrak{F}_2 \qquad \mathfrak{F}_1{}^2 + II'_T/3\cdot\mathfrak{F}_2{}^2 \end{vmatrix}}$$

$$F_2 = \frac{\mathfrak{F}_2}{|\ \text{as above}\ |}.$$

The moduli of elasticity F are therefore generally not the reciprocals of the coefficients of elasticity $\mathfrak{F}$.

We now carry out in imagination a series of experiments such as mentioned at the end of Section **2**.

(i) Firstly, we apply a uniform hydrostatic pressure; here the stress tensor is a scalar tensor

$$(8.6)\qquad T_s{}^r = -p\delta_s{}^r$$

where p is what is commonly called "pressure" and the stress invariants are

$$(8.7)\qquad T = -p, \qquad II'_T = III'_T = 0$$

$T'_s{}^r$ and $T'_a{}^r T'_s{}^a$, therefore, vanish and the second of (8. 4) gives $\epsilon'_s{}^r = 0$, while the first yields

$$(8.8)\qquad \epsilon = -p\mathfrak{F}_{01}(T, 0, 0).$$

This defines a *coefficient of volume elasticity*

$$(8.9)\qquad k' = -3\epsilon/p = 3\mathfrak{F}_{01}.$$

[7] For the derivation compare Reiner.

Considering that II'_ϵ and III'_ϵ vanish, the first of (8.3) gives

$$T = -p = \epsilon \mathfrak{F}_{01}. \tag{8.10}$$

This defines the *modulus of compression*

$$k = -p/3\epsilon = F_{01}/3 \tag{8.11}$$

and $k' = 1/k$.

(ii) In the second experiment we apply a tangential stress

$$T_s{}^r = \begin{Vmatrix} 0 & T_{xy} & 0 \\ T_{xy} & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix} = T'_s{}^r \tag{8.12}$$

so that

$$T = 0; \qquad II'_T = -T_{xy}{}^2; \qquad III'_T = 0 \tag{8.13}$$

and

$$T'_a{}^r T'_s{}^a = T_{xy}{}^2 \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{Vmatrix}. \tag{8.14}$$

This makes (8.4)

$$\epsilon = -T_{xy}{}^2 \mathfrak{F}_{02}(0, II'_T, 0)$$

$$\epsilon'_s{}^r = \mathfrak{F}_1(0, II'_T, 0) T_{xy} \begin{Vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{Vmatrix} + \mathfrak{F}_2(0, II'_T, 0) \frac{T_{xy}{}^2}{3} \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{Vmatrix} \tag{8.15}$$

and defines three coefficients of elasticity, *viz.*

$$\begin{aligned} \delta' &= -\epsilon/II'_T = -\mathfrak{F}_{02} \\ \mu' &= 2\mathfrak{F}_1 \\ \alpha' &= -2\mathfrak{F}_2/3. \end{aligned} \tag{8.16}$$

The coefficient $\mathfrak{F}_1$ connects shearing stress with shearing strain and is accordingly a generalized *coefficient of shear elasticity* or *of rigidity*. The isotropic component of the strain, ϵ, is a measure of the cubical dilation. If δ' does not vanish, a simple shearing stress will produce an increase (or decrease for negative δ') of the volume measured by $\delta' T_{xy}{}^2$.[8]

[8] It is remarkable that Sir William Thomson (Lord Kelvin) should have foreseen in 1875 the possible existence of such a phenomenon on purely theoretical grounds, *vide* the following quotation: "It is possible that a shearing stress may produce in a truly isotropic solid condensation or dilatation in proportion to the square of its value; and it is possible that such effect may be sensible in india-rubber or cork, or other bodies susceptible of great deformations or compressions with persistent elasticity." Footnote p. 34, *Math. & Phys. Papers*, Vol. III, London, 1890. Weissenberg has observed negative elastic dilatancy in porous rubber (not yet published).

Accordingly, δ' may be termed the *coefficient of* (elastic) *dilatancy* (compare Reiner). Should δ' vanish, but not α', then a simple shearing stress will produce (in the case of a positive α') an extension normal to the plane of shear (in our case the z-direction) which is equal to $\alpha' T_{xy}{}^2$, together with two lateral contractions equal to $\alpha'/2 \cdot T_{xy}{}^2$, so that the volume is not changed. If δ' should not vanish, there will be superposed a change of volume. We may call α' the *coefficient of cross-elasticity.*

(iii) If we force upon the material a tangential strain, we shall similarly find three moduli of elasticity

$$\begin{aligned} \delta &= -F_{02}/4 \\ \mu &= F_1/2 \\ \alpha &= -F_2/6 \end{aligned} \tag{8.17}$$ [9]

of which μ is a generalized *shear modulus* or *modulus of rigidity.* δ, the *modulus of dilatancy,* will measure a hydrostatic tension necessary to maintain simple shear; and α, a *modulus of cross-elasticity,* measures a stress produced by simple shear, in the direction normal to its plane.

(iv) Simple pull, in infinitesimal elasticity employed to determine Young's modulus and Poisson's ratio, gives us

$$T_s{}^r = \left\| \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & T_{zz} \end{matrix} \right\| . \tag{8.18}$$

so that

$$T = T_{zz}/3 \,; \; T'_s{}^r = \frac{T_{zz}}{3} \left\| \begin{matrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{matrix} \right\| ; \; T'_a{}^r T'_s{}^a = \frac{T_{zz}{}^2}{9} \left\| \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{matrix} \right\| \tag{8.19}$$

and.

$$II'_T = -T_{zz}{}^2/3 \,; \qquad III'_T = 2T_{zz}{}^3/27. \tag{8.20}$$

This makes

$$\begin{aligned} \epsilon_s{}^r = T_{zz}/3\,(k'/3 + T_{zz}\delta' + (2T_{zz}{}^2/9)\,\mathfrak{F}_{03}) \left\| \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \right\| \\ + (T_{zz}/6)\,(\mu' - \alpha' T_{zz}) \left\| \begin{matrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{matrix} \right\| \end{aligned} \tag{8.21}$$

---

[9] Note that a simple shear is measured traditionally by twice the tangential component of the strain-tensor.

and defines a generalized Youngs' modulus

$$(8.22)\quad E^{-1} = \epsilon_{zz}/T_{zz} = -\tfrac{1}{3}[k'/3 + \mu' + T_{zz}(\delta' - \alpha') + (2T_{zz}^2/9)\mathfrak{F}_{03}]$$

and a generalized Poisson-ratio

$$(8.23)\quad \sigma = -\epsilon_{xx}/\epsilon_{zz} = -\frac{k'/3 - \mu'/2 + T_{zz}(\delta' + \alpha'/2) + (2T_{zz}^2/9)\mathfrak{F}_{03}}{k'/3 + \mu' + T_{zz}(\delta' - \alpha') + (2T_{zz}^2/9)\mathfrak{F}_{03}}.$$

Either E or σ can be used to determine a further coefficient of elasticity

$$(8.24)\qquad \beta' = 2\mathfrak{F}_{03}/9.$$

Summarizing, we can now write (8.4) as follows:

$$(8.26)\qquad \begin{aligned} \epsilon_v &= -k'p - 3\delta' II'_T + (27/2)\beta' III'_T \\ \epsilon'_s{}^r &= (\mu'/2)T'_s{}^r - 3(\alpha'/2)(T'_a{}^r T'_s{}^a + 2II'_T/3 \cdot \delta_s{}^r) \end{aligned}$$

and (8.3) as follows:

$$(8.27)\qquad \begin{aligned} p &= -k\epsilon_v + 4\delta II'_\epsilon - (9\beta/2) III'_\epsilon \\ T'_s{}^r &= 2\mu\epsilon'_s{}^r - 6\alpha(\epsilon'_a{}^r \epsilon'_s{}^a + 2II'_\epsilon/3 \cdot \delta_s{}^r) \end{aligned}$$

where p is the hydrostatic pressure and ϵ_v the cubical dilatation, $\epsilon'_s{}^r$ the deviator of strain and $T'_s{}^r$ the deviator of stress, k, δ, β, μ, α moduli of elasticity and k', δ', β', μ', α' coefficients of elasticity. These are generally functions of all three invariants of stress and strain respectively, but may also degenerate to constants. A hydrostatic tension will cause a cubical dilation and *vice versa*; but a cubical dilatation may also be caused in the absence of a hydrostatic tension by either simple shearing stress or traction. Likewise, a hydrostatic pressure may be required to maintain simple shear or a volume-constant simple extension. Finally, a simple shearing stress may not only produce a corresponding shearing strain, but also "sideways" a volume-constant extension. Likewise simple shear may require for its maintenance not only a corresponding shearing stress but also "sideways" a traction. The general elastic body has accordingly three additional properties absent in classical elasticity, namely dilatancy of two kinds, (shear- and tractional dilatancy) and cross-elasticity. It is not so much the property of dilatancy predicted by Kelvin as early as 1875 and observed as a permanent set by Reynolds as early as 1885, which is challenging, but the cross-elasticity, which is connected with the functions $\mathfrak{F}_2$ and F_2 respectively. We, therefore, consider this property again from a different aspect.

9. Let n be the normal to an element of interface in the interior or of surface on the boundary of the body under consideration. Let the traction

T_n be resolved into three orthogonal components T_{nq} where q runs through n, t and c; t being the direction parallel to the face and c the direction crosswise to n and t, so that

$$(9.1) \qquad T_{nc} = 0.$$

Let ϵ_n be resolved in the same directions. We find, then, from the second of (8.4)

$$(9.2) \qquad \epsilon'_{nc} = \mathcal{F}_2 T'_{na} T'_{ac}$$

the term following within brackets disappearing because $r \neq s\,(n \neq c)$. Now

$$(9.3) \qquad T'_{na}T'_{ac} = T'_{nn}T'_{nc} + T'_{nt}T'_{tc} + T'_{nc}T'_{cc}.$$

As (9.2) is not affected by an isotropic stress component, T'_{nc} vanishes also and this reduces (9.3) to

$$(9.4) \qquad T'_{na}T'_{ac} = T'_{nt}T'_{tc}.$$

Now on the right side of (9.4) T'_{nt} does not vanish, by definition; and if one imagines in the standard cube which defines T_{xx} etc., x, y, z replaced by n, t, c, it is clear that T'_{tc} will, in general, not vanish. Therefore ϵ'_{nc} is finite. This brings out very strikingly a consequence of the existence of $\mathcal{F}_2$ and supports the designation "cross-elasticity." We have, however, shown that $\mathcal{F}_2$ (or F_2) can generally not be omitted without prejudicing experimental results.

TEL AVIV, PALESTINE.

BIBLIOGRAPHY.

Hencky, H., *Annalen der Physik* (5), vol. 2 (1929), p. 617.
Love, A. E. H., *Mathematical Theory of Elasticity*, Cambridge, 1906.
Murnaghan, F. D., *American Journal of Mathematics*, vol. 59 (1937), p. 235.
Prager, W., *Journal of Applied Physics*, vol. 16 (1945), pp. 837-840.
Reiner, M., *American Journal of Mathematics*, vol. 67 (1945), p. 35.
Roentgen, *Annalen der Physik*, vol. 159 (1876), p. 601.
Sokolnikoff, I. S., *Mathematical Theory of Elasticity*, New York, 1946.
Stokes, G. G., *Transactions of the Cambridge Philosophical Society*, vol. 8 (1849), p. 287.
Taylor, G. I. and Quinney, H., *Proceedings of the Royal Society*, A., vol. 143 (1934), p. 307.
Wall, F. T., *Journal of Chemical Physics*, vol. 10 (1942), p. 132.
Weissenberg, K., *Akademie der Wissenschaften*, Berlin, 1931.
Weissenberg, K., *British Rheologist's Club Conference*, London, 1946.
Weissenberg, K., *Nature*, vol. 159 (1947), p. 310.

LOCALLY COMPACT RINGS.*

By Irving Kaplansky.

1. Introduction. In this paper an account is given of some results in the structure theory of locally compact (= bicompact) rings. We begin by recapitulating, in somewhat generalized form, the known results on locally compact connected rings (Theorem 1). The theory of locally compact rings without nilpotent ideals is shown to be reducible to the totally disconnected case (Theorem 2). In **3** the hypothesis of boundedness is added, and a complete result given for the semi-simple case (Theorem 4). This section concludes with remarks on a class of rings including both compact rings and discrete rings with descending chain condition; it is shown that the known structure theorems for these two cases can thus be unified. The next two sections are devoted to maximal ideals, the existence and continuity of inverses, and the effect of chain conditions. A principal tool in this investigation is the fact (Lemma 4) that a locally compact totally disconnected ring has compact open subrings; this makes it possible to apply the structure theory for compact rings given in [11]. The final section assembles some results on locally compact primitive rings.

2. The component of 0. We shall use the definition of boundedness in [11]: a subset S of a topological ring A is *right bounded* if for any neighborhood U of 0 there exists a neighborhood V such that $V \cdot S \subset U$ (by $V \cdot S$ we mean the set of all products of elements of V by elements of S). Boundedness means both right and left boundedness, the latter being analogously defined.

The following theorem contains various known results as special cases; cf. [2], [5], [8], [11, Th. 8], [14], [15]. The proof is virtually the same as that in [11] but we repeat it for completeness.

Theorem 1. *If A is a locally compact ring with C the component of* 0, *and B is a right bounded subgroup of A, then $CB = 0$.*

Proof. For any fixed character f of A, let $I(f)$ denote the set of all $a \varepsilon A$ with $f(aB) = 0$. Then $I(f)$ is clearly a subgroup of A; if we show it to be open it will contain C and we will have $f(CB) = 0$ for all f, so that

* Received August 12, 1947.

$CB = 0$. Choose neighborhoods U, V of 0 in A with $f(U) < 1/2$, $V \cdot B \subset U$. Then for $x \varepsilon V$ we have $nxB \subset U$ for any integer n, and hence $f(nxb) = nf(xb) < 1/2$ for all $b \varepsilon B$. Hence $f(xb) = 0$, $V \subset I(f)$, $I(f)$ is open.

We now quote the structure theorem [18, p. 110] which asserts that A is the direct sum (in the topological group sense) of a vector group N and a group in which the component P of 0 is compact. Since P is compact it is bounded [11, Lemma 10]. Hence we have the following corollary.

COROLLARY. *With the above notation* $P^2 = PN = NP = 0$.

Before proceeding to the next theorem, we make a remark on direct sums. If A is a topological ring and B, C are closed ideals with $B \cap C = 0$, $B + C = A$, then A is the direct sum of B and C in the ring-theoretic but not necessarily in the topological sense. However in a more special case we are able to assert that we have a direct sum in both senses: if B has a unit element e. For if the directed set a_i approaches a, then ea_i approaches ea, i.e., the B-components of a_i approach the B-component of a. By subtraction we get the same result for the C-components, and this verifies that we have the Cartesian product topology in $B + C$.

THEOREM 2. *A locally compact ring with no algebraically nilpotent* [1] *ideals is the direct sum of a connected ring and a totally disconnected ring. The former is a semi-simple algebra of finite order over the real numbers.*

Proof. In the notation of the Corollary to Theorem 1 we must have $P = 0$, since P is a nilpotent ideal. Then N becomes the component of 0 and is an ideal. Any nilpotent ideal in N would generate a nilpotent ideal in A; hence N has no nilpotent ideals. Also N is not only a vector space but an algebra over the real numbers; this requires only the verification of the appropriate associativity condition [cf. 8]. Thus N is a semi-simple algebra over the reals and has a unit element. We form the Peirce decomposition with respect to the latter, and we have $A = N + S$ where $S = A - N$ is totally disconnected.

3. Bounded rings. We begin with certain remarks valid for any bounded rings.

[1] A set S is algebraically nilpotent if for some n, $S^n = 0$. We have prefixed the term "algebraic" in order to distinguish between this and the topological nilpotence defined on p. 162 of [11].

LEMMA 1. *In a bounded ring the quasi-inverse* [2] *is uniformly continuous.*

Proof. We use a' to denote the quasi-inverse of a. The identity

$$a' - b' = (1 + a')(b - a)(1 + b')$$

shows that in a bounded ring, $a' - b'$ can be made arbitrarily small by choosing $a - b$ sufficiently small.

THEOREM 3. *In a complete bounded ring, the quasi-regular elements form a closed set.*

Proof. Suppose the directed set a_i converges to a and that a'_i exists for every i. Lemma 1 shows that a'_i is a Cauchy directed set. Its limit a' is the quasi-inverse of a.

From [11, Th. 4] we derive the following corollary.

COROLLARY. *The radical of a complete bounded ring is closed.*

The following lemma gives a new proof of Jacobson's result [10, Th. 26] that a two-sided ideal in a semi-simple ring is semi-simple.

LEMMA 2. *If B is a two-sided ideal in a ring A, we have $R(B) = R(A) \cap B$, where R denotes the radical in each case.*

Proof. First on the mere assumption that B is a right ideal, we prove $R(B) \supset R(A) \cap B$. Let $x \varepsilon R(A) \cap B$; then x has a quasi-inverse y which is in B since $y = -x - xy$. Hence $R(A) \cap B$ consists of elements which are quasi-regular in B, and since it is an ideal in B, it is contained in $R(B)$. Conversely suppose $x \varepsilon R(B)$, where B is now a two-sided ideal in A. Then for any $a \varepsilon A$, $-(xa)^2$ is in $R(B)$ and is quasi-regular, say with quasi-inverse z. Hence $-xa$ has the quasi-inverse $xa o z$, and $x \varepsilon R(A)$.

The following structure theorem reduces the study of locally compact bounded semi-simple rings to the discrete case, since compact semi-simple rings are completely known [11, Th. 16].

THEOREM 4. *A locally compact bounded semi-simple ring is the direct sum of a compact semi-simple ring and a discrete semi-simple ring.*

Proof. It follows from Theorem 1 that A is totally disconnected. Since A is bounded and locally compact, it has a system of ideal neighborhoods of 0 [11, Lemma 9]. In particular we have a compact open ideal B, which

[2] The terms quasi-inverse, radical, semi-simple, and primitive are used as defined by Jacobson in [10]. We also use the notation $x o y = x + y + xy$.

is semi-simple by Lemma 2. B has a unit element [11, Th. 16], and we may use the Peirce decomposition to write $A = B + C$, with $C = A - B$ discrete since B is open. That C is semi-simple follows from another application of Lemma 2.

The presence of a compact open ideal may be used to obtain results without the assumption of semi-simplicity. The following is an example.

THEOREM 5. *A commutative locally compact totally disconnected bounded ring is the direct sum of a compact ring with unit, and a ring which modulo its radical is discrete.*

Proof. Our assumptions imply the existence of a compact open ideal I, which in turn [11, Th. 17] is the direct sum of a compact ring B with unit e, and a radical ring D. We form the Peirce decomposition $A = B + C$, with $B = Ae$ and C the annihilator of e. Now D is an ideal in I, which is an ideal in A. By two applications of Lemma 2 we have $R(D) = R(A) \cap D$, i. e. $D \subset R(A)$. Since $R(A) = R(B) + R(C)$ and $D \subset C$, we have $D \subset R(C)$. Now $C - D = (B + C) - (B + D) = A - I$ is discrete. Hence $C - R(C)$ is discrete.

We shall conclude this section with some remarks on a still more special class of rings. These rings satisfy the following three conditions: (1) local compactness, (2) boundedness, (3) the descending chain condition for right ideals which contain a fixed open two-sided ideal. Special cases are compact rings and discrete rings with the descending chain condition on right ideals. We shall now briefly indicate how the classical results in the latter case, and the results for compact rings in [11], may be thus unified and generalized.

First we consider the semi-simple case. Theorem 4 is applicable and condition (3) shows that the discrete summand has the descending chain condition, and accordingly is a direct sum of a finite number of matrix rings over division rings. The result may be restated thus: a semi-simple ring satisfying (1), (2), and (3) is the Cartesian direct sum of matrix rings over division rings, all but a finite number of the components being finite.

In the general case let R be the radical of a ring A satisfying (1), (2) and (3). By the Corollary to Theorem 3, R is closed. The arguments of [11, p. 163] may now be repeated virtually verbatim, and we have that R is the union of all (topologically) nil left and right ideals, and in the totally disconnected case is itself topologically nilpotent. The idempotents of $A - R$ may be transferred to A as in [11, Lemma 12] and this yields the analogues of Theorems 17 and 18: if A is commutative it is the direct sum of primary

rings and a radical ring, and a primary ring is the ring of matrices over a completely primary ring.

4. Quasi-inverses. We repeat from [11] the following definition: a topological ring is a Q_r-ring if the right quasi-regular elements form an open set, a Q-ring if the quasi-regular elements form an open set. It suffices [11, Lemma 2] that there be a neighborhood of 0 consisting of (right) quasi-regular elements. Moreover [11, Lemma 4] in a Q-ring, the quasi-inverse is continuous wherever defined if it is continuous at 0.

LEMMA 3. *Let B be a closed two-sided ideal in the topological ring A. Then A is a Q-ring with continuous quasi-inverse if and only if both B and $A-B$ are.*

Proof. Suppose that A is a Q-ring with continuous quasi-inverse. If x is near 0 in B, then x has a quasi-inverse $y=-x-xy$ necessarily in B, which, being near 0 in A is also near 0 in B. Again, for any element z near 0 in $A-B$, we may pick an inverse image in A near 0 in A. The image of the latter's quasi-inverse provides us with a quasi-inverse of z near 0 in $A-B$.

Conversely suppose that both B and $A-B$ are Q-rings with continuous quasi-inverse. For x near 0 in A the coset $x+B$ is near 0 in $A-B$. The quasi-inverse of $x+B$ is a coset having a representative y near 0 in A. Then xoy is in B, and being small, has a small quasi-inverse z. Then yoz is the desired quasi-inverse of x.

We shall now apply Lemma 3 to locally compact rings. First we dispose of the connected case.

THEOREM 6. *A connected locally compact ring is a Q-ring with continuous quasi-inverse.*

Proof. In the notation of the Corollary to Theorem 1, P is a closed two-sided ideal which is certainly a Q-ring with continuous quasi-inverse in virtue of $P^2=0$. Next $A-P$ is an algebra of finite order over the reals. It can be normed [1, Th. 4] and hence [11, Lemmas 3, 5] is a Q-ring with continuous quasi-inverse. The result now follows from Lemma 3.

The treatment of the totally disconnected case is based upon the following lemma.

LEMMA 4. *A locally compact totally disconnected ring A has a system of neighborhoods of 0 which are compact open subrings.*

Proof. We know that A has subgroup neighborhoods of 0. Let U be such a compact open subgroup, and select an open subgroup V such that $V \subset U$, $VU \subset U$ (this is possible since U is bounded). Define $W = V + V^2 + V^3 + \cdots$. Then W is a compact open subring of A and $W \subset U$.

For later use we note the following refinement.

LEMMA 5. *Let A be a locally compact totally disconnected ring having a unit element* 1 *such that the closed subring C generated by* 1 *is compact. Then A has a compact open subring containing* 1.

Proof. If U is any compact open subring, the desired subring is $U + C$.

We remark that the hypothesis of Lemma 5 is of course satisfied if A has finite characteristic. Another case where the hypothesis holds is where A has a compact open subgroup U containing an element a such that the mapping $x \to xa$ is a homeomorphism (for example, A may be any non-discrete locally compact totally disconnected division ring); for then C is homeomorphic to Ca which is compact. An example where the hypothesis fails is the ring of integers under the discrete topology.

LEMMA 6. *A compact ring A which is not a Q-ring contains a set of non-zero idempotents with cluster point* 0.

Proof. If the radical R of A is open, A is clearly a Q-ring. Hence we suppose that R is not open, which means that $A - R$ is infinite. By [11, Th. 16] we may find in $A - R$ an infinite set of idempotents having only 0 as a cluster point. We may build [11, Lemma 12] idempotents e_i in A mapping on these. The set e_i has a cluster point e, necessarily an idempotent, and necessarily in R since it maps on 0 mod R. But the negative of an idempotent is right quasi-regular only if it is 0, and so $e = 0$.

We now prove the principal result of this section. It provides for the locally compact case an affirmative answer to the question raised in [11] as to whether any Q_r-ring is a Q-ring.

THEOREM 7. *A locally compact ring A is a Q-ring with continuous quasi-inverse if and only if A is a Q_r-ring.*

Proof. In the light of Lemma 3 and Theorem 6, it will suffice to treat the totally disconnected case. Then by Lemma 4, A has a compact open subring B. We assert that B is a Q-ring, for otherwise by Lemma 6 there are negatives of non-zero idempotents arbitrarily near 0, which is impossble

in a Q_r-ring. It then follows that A is a Q-ring. That the quasi-inverse is continuous is a consequence of Lemma 1.

We may now prove Theorems 21 and 22 of [11] without countability assumptions.

THEOREM 8. *A locally compact ring without divisors of* 0 *is a Q-ring with continuous quasi-inverse.*

Proof. The proof is like that of Theorem 7, except that the presence of idempotents near 0 is now ruled out by the remark that the only possible idempotents are 0 and 1.

We prove one further result of this kind.

THEOREM 9. *A locally compact ring with the ascending or descending chain condition on closed right ideals is a Q-ring with continuous quasi-inverse.*

Proof. Assuming that the usual compact open subring is not a Q-ring we again find an infinite set of idempotents; but this time we arrange, as is clearly possible in the light of [11, Th. 16], that $e_ie_j = e_je_i = e_j$ for $i \geqq j$ or $i \leqq j$ according as we have the ascending or descending chain condition. In the former case $\{e_iA\}$ is an ascending set of closed right ideals. But if $e_iA = e_jA$ for $i > j$ then $e_i = e_ja$, and left-multiplying by e_j gives $e_j = e_ja = e_i$, a contradiction. The argument is similar for the descending case.

In Theorems 7, 8 and 9 we have generalized Otobe's result [16] that the inverse is continuous in a locally compact division ring. These do not exhaust the possible hypotheses that are adequate. We mention one more: assume no nilpotent ideals (so that Theorem 2 is applicable), and assume outright that there is a neighborhood of 0 free of non-zero idempotents.

At the end of **5** we shall give an example of a locally compact ring in which the quasi-inverse is not continuous.

We conclude this section with the following supplement to Theorem 9.

THEOREM 10. *A semi-simple Q-ring with the descending chain condition on closed right ideals has the descending chain condition on all right ideals. Hence a locally compact semi-simple ring with the descending chain condition on closed right ideals is the direct sum of a finite number of matrix rings over locally compact division rings.*

Proof. We shall use Jacobson's structure theory of semi-simple rings

[10, pp. 310-312], sharpened by the use of Segal's notion of regular ideals.[3] We may summarize the facts that we need as follows: the regular maximal right ideals M_i in A have intersection 0. We form for each i the ideal $P_i = (M_i : A) =$ the set of all $x \varepsilon A$ with $Ax \subset M_i$. By the use of right multiplications the primitive ring $A - P_i = Q_i$ may be represented as a dense ring of endomorphisms in the vector space $A - M_i$.

Now in our topological context we note further that M_i is closed [17, Th. 1.6], and from this it follows readily that P_i is closed. Let $x_1, x_2, \cdots$ be linearly independent elements in the vector space $A - M_i$ in which Q_i acts, and let I_n denote the set of elements (i.e. linear transformations) annihilating $x_1, \cdots, x_n$. It is easy to see that the I's form a properly descending chain of closed right ideals. Since this chain must terminate $A - M_i$ is finite-dimensional. Next we note that a finite number of the M's already have intersection 0. From this we can conclude that A is the direct sum of a finite number of simple rings (*cf.* the argument at the foot of p. 314 of [10]). In the locally compact case we use Theorem 9 to get that A is a Q-ring, and thus verify the second sentence of the theorem. We remark finally that the direct sum and matrix representations involved hold in the topological as well as in the algebraic sense, as can be seen by persistent use of idempotents.

5. Maximal ideals. The results in this section will be obtained by further exploitation of the existence of compact open subrings. We begin with two preliminary results.

LEMMA 7. *Let e be an idempotent in a topological ring A, such that eAe is a Q_r-ring. Then any maximal right ideal M containing $\{a - ea\}$ is closed.*

Proof. Suppose on the contrary that the closure of M is A. Then M contains $e + x$ with x arbitrarily small, say so small that exe is right quasi-regular, $exe \circ y = 0$. Then M contains $z = (e + x)(e + ey)$, also $ez - z$, and hence ez. A computation shows that $ez = e$, so that M contains e and all of A, a contradiction.

LEMMA 8. *Let e be an idempotent in a topological ring A, such that eAe is a Q_r-ring. Then if x lies in every closed regular maximal right ideal of A, ex is in the radical of A.*

---

[3] A right ideal I is regular if there exists a left unit e modulo I, that is, an element e such that $ex - x \varepsilon I$ for all x.

Proof. Consider the right ideal I generated by $\{a - ea\}$ and $e + xe$. If I is a proper ideal it can be expanded to a maximal right ideal M excluding e, which is closed by Lemma 7, and regular since e is a left unit. Then $x \varepsilon M$, $xe \varepsilon M$, $e \varepsilon M$, $M = A$, a contradiction. Hence $I = A$, and we must have

$$b + (e + xe)(e + d) = e$$

where $b \varepsilon \{a - ea\}$, $d \varepsilon A$. Left-multiplying by e we find $exe \circ ed = 0$. Thus $(ex)e$ is right quasi-regular and hence so is $e(ex) = ex$ (cf. the last identity on p. 154 of [11]). Since xc is also in every closed regular maximal right ideal for every c, we likewise have exc right quasi-regular, and hence ex is in the radical.

THEOREM 11. *A locally compact ring which is not a radical ring has a closed regular maximal right ideal.*[4]

Proof. First we consider the totally disconnected case. We find a compact open subring B. If B is a Q-ring, so is A, and (unless A is a radical ring) we know that A has regular maximal right ideals which are closed. Otherwise, let R be the radical of B. In $B - R$ we select a primitive idempotent, and we find in B an idempotent e mapping on f. Now we know that f annihilates a neighborhood of 0 in $B - R$. Hence for a suitable neighborhood U of 0 in B we have $eUe \subset R$. This shows that eAe is a Q-ring, and Lemma 7 then provides us with the desired closed regular maximal right ideal, namely any right ideal containing $\{x - ex\}$ and maximal with respect to exclusion of e.

In the general case let C be the component of 0 in A. If $A - C$ is a radical ring it is a Q-ring, so is A (Lemma 3 and Theorem 6), and we are finished. Otherwise by the first part of the proof $A - C$ has a closed regular maximal right ideal. We take its inverse image in A.

THEOREM 12. *Let A be a locally compact totally disconnected ring with a unit element* 1 *such that the closed subring generated by* 1 *is compact. Then the intersection S of the closed regular maximal right ideals in A is the closure R' of the radical R of A.*

Proof. That S contains R' is clear, since S is closed and contains R. Conversely suppose $x \varepsilon S$. We select (Lemma 5) a compact open subring B containing 1; let T be the radical of B. In $B - T$ we may find a directed set f_i of idempotents of finite rank approaching 1 (an idempotent of finite

---

[4] Concerning this question of the existence of closed maximal ideals, R. Arens has remarked that his ring L^{ω} [3] has no closed maximal ideals at all.

rank is the sum of a finite number of primitive idempotents); this is an immediate consequence of [11, Th. 16]. We build idempotents e_i in B mapping on f_i [11, Lemma 12]. Then $\{e_i\}$ has the cluster point 1, for such a cluster point is necessarily an idempotent mapping on the unit element of $B - T$, and 1 is the only element fulfilling these conditions. By Lemma 8, $e_i x \varepsilon R$ and it follows that $x \varepsilon R'$.

The significance of a result like Theorem 12 is as follows. Suppose that A is a locally compact semi-simple ring. Then $R = R' = 0$, and under the hypothesis of Theorem 12, $S = 0$. We then immediately deduce a representation of A as a subdirect sum of locally compact primitive rings, the point being that only with *closed* maximal ideals is a topology inherited by the components of the sum. It need hardly be added that if A is a Q-ring, the assumption about the unit element is superfluous. Whether it is in any event superfluous I have not been able to determine.

We give an example now of the kind of subdirect sum that can arise in locally compact rings.[5] Let F_p denote the p-adic numbers, I_p the p-adic integers. Let A be the set of all sequences $\{a_i\}$ of elements of F_p, with all but a finite number of components in I_p; let B be the subset with all components in I_p. The topology is as follows: neighborhoods of 0 in B are taken to be neighborhoods of 0 in A, and B is given the Cartesian product topology. Then B is compact so that A is locally compact. Continuity of addition and multiplication in A are readily verified. A is semi-simple: in fact the sequences with 0 at a designated place form a closed maximal ideal, and these ideals have intersection zero. A noteworthy feature of this ring is that the (multiplicative) inverse is not continuous. The elements b_i with p in the i-th place and 1 elsewhere approach the unit element of A (the sequence of all 1's), but b_i^{-1} does not approach the unit element since the difference does not even enter B.

6. Primitive rings. As was observed earlier, Theorem 12 provides a partial reduction of locally compact semi-simple rings to the primitive case. Thus an important step in the theory would be accomplished if locally compact primitive rings were classified. It follows from Theorem 2 and [10, Lemma 4] that only the totally disconnected case need be studied.

The most important result so far obtained is the structure theorem of Jacobson [9]: any non-discrete locally compact division ring admits a valuation and is an algebra of finite order over its center. There have been three

[5] This kind of construction is closely related to certain group-theoretic results of Braconnier and Dieudonné [7].

later contributions: Otobe [16] removed the countability assumption in Jacobson's proof, Braconnier [4] announced a proof based on Haar measure, and a new proof is given in [12]. The latter proof leans strongly on the following result.

LEMMA 9. *A locally compact topological linear space over a non-discrete locally compact division ring is finite-dimensional.*

The connected case is a classical result of F. Riesz. The general case was announced in [6] and a proof is given in [12]. In the characteristic zero case a proof was already given by Jacobson in [9], and we wish to record here that the characteristic zero case also follows from a theorem of Mackey [13].

We turn now from division rings to primitive rings. Jacobson [9, Th. 5.3] gave the following result: a non-discrete locally compact simple [6] ring S of characteristic zero is an algebra of finite order over the p-adic numbers. However there appears to be a gap in the proof that is not easy to supply: the tacit assumption (especially on p. 440) that multiplication by p^{-1} is continuous, so that p^nG is open where G is a compact open subgroup of S. The more general statement is made on p. 442 that if S is a locally compact p-group [7] which is an algebra over the rational numbers, then S is an algebra of finite order over the p-adic numbers. That this statement is in error is shown by the following example: let S be the set of all sequences of p-adic numbers with the proviso that in each sequence there is a bound to the negative powers of p. Let T be the subset of S consisting of those sequences with all components p-adic integers. We give T the Cartesian product topology and make it open in S. Multiplication is ruled out of the picture by setting $S^2 = 0$. Then S is locally compact and a p-group, but it is infinite-dimensional over the p-adic numbers. It goes without saying that multiplication by p^{-1} is not continuous, and consequently S is not a *topological* linear space over the p-adic numbers.

If we add to the hypothesis of simplicity the existence of a unit, then p^{-1} becomes a member of S and multiplication by it will be continuous. However a somewhat shorter proof is made possible by the assumption of a unit.

THEOREM 13. *A non-discrete locally compact simple ring A of characteristic zero with unit element is an algebra of finite order over its center.*

We first note the following lemma.

---

[6] A simple ring here means one with no proper ideals, and not merely no proper closed ideals.

[7] A p-group is one in which every element satisfies $p^n a \to 0$.

LEMMA 10. *For any compact commutative totally disconnected group G, there exists a prime p and an element $a \neq 0$ in G such that $p^n a \to 0$.*

The assumption of the first axiom of countability in Jacobson's proof [9, Lemma 4.4] can be avoided without difficulty. A brief proof can also be based on character theory: the character group H of G is a discrete group with all its elements of finite order. The elements of H with order a power of p form a direct summand H_p which must be non-zero for some p. The annihilator G_p of $H - H_p$ is easily seen to be a p-group.

Proof of Theorem 13. The center of any simple ring with unit is a field. Hence the center Z of A is a locally compact field. If we show that Z is non-discrete, then by Lemma 9 $[A:Z]$ will be finite. Now for any p the set of elements a in A such that $p^n a \to 0$ forms an ideal A_p. By Lemma 10 this ideal is non-void for some p, and $A_p = A$ since A is simple. In particular $p^n \cdot 1 \to 0$, Z is non-discrete.

We present another result that can be proved in a similar way.

THEOREM 14. *Let A be a non-discrete locally compact primitive ring and suppose that A is of characteristic zero and has minimal ideals. Then A is an algebra of finite order over its center.*

Proof. Let eA be a minimal right ideal. Then eA is a locally compact topological linear space over the locally compact division ring eAe, and A is represented as a dense ring of endomorphisms on this linear space. Again we have only to prove that eAe is non-discrete, for this will make eA finite-dimensional. By Lemma 10 we have a non-zero ideal I consisting of all a with $p^n a \to 0$. By [10, Th. 29], e is in I. Hence eAe is not discrete.

We shall conclude with an example showing that in the characteristic p case, it is possible for a locally compact primitive ring to be infinite-dimensional over its center. Let K be a finite field and A the set of all infinite matrices $a = (a_{ij})$, $a_{ij} \in K$, which are "ultimately triangular": for any a there exists $N = N(a)$ such that $a_{ij} = 0$ for $j > i > N$. Let B be the subset of matrices which are actually triangular: $a_{ij} = 0$ for $j > i$. We topologize B with the weak topology, a general neighborhood of 0 in B consisting of all matrices with the first n rows zero. This makes B compact, and by declaring that B is open in A we make A locally compact. That A is primitive is clear and incidentally it has a unit element and minimal ideals. The one point needing serious verification is the continuity of multiplication, a verification which we omit.

This ring is not simple: the linear transformations with finite-dimen-

sional range form a proper ideal in A. However A is simple in the sense that there are no proper *closed* ideals. It seems unlikely that a locally compact ring with minimal ideals could be simple in the strong sense, since completeness appears to require the presence of linear transformations with infinite-dimensional range.

UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY.

1. A. A. Albert, "Absolute valued real algebras," *Annals of Mathematics*, vol. 48 (1947), pp. 495-501.

2. H. Anzai, "On compact topological rings," *Proceedings of the Imperial Academy of Tokyo*, vol. 19 (1943), pp. 613-615.

3. R. Arens, "The space L^ω and convex topological rings," *Bulletin of the American Mathematical Society*, vol. 52 (1946), pp. 931-935.

4. J. Braconnier, "Groupes d'automorphismes d'un groupe localement compact," *Comptes Rendus de l'Académie des Sciences, Paris*, vol. 220 (1945), pp. 382-384.

5. J. Braconnier, "Sur les modules localements compacts," *Comptes Rendus de l'Académie des Sciences, Paris*, vol. 222 (1946), pp. 527-529.

6. J. Braconnier, "Sur les espaces vectoriels localements compacts," *Comptes Rendus de l'Académie des Sciences, Paris*, vol. 222 (1946), pp. 777-778.

7. J. Braconnier and J. Dieudonné, "Sur les groupes abeliens localements compacts," *Comptes Rendus de l'Académie des Sciences, Paris*, vol. 218 (1944), pp. 577-579.

8. N. Jacobson and O. Taussky, "Locally compact rings," *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 106-108.

9. N. Jacobson, "Totally disconnected locally compact rings," *American Journal of Mathematics*, vol. 58 (1936), pp. 433-449.

10. N. Jacobson, "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

11. I. Kaplansky, "Topological rings," *American Journal of Mathematics*, vol. 69 (1947), pp. 153-183.

12. I. Kaplansky, "Topological methods in valuation theory," *Duke Mathematical Journal*, vol. 14 (1947), pp. 527-541.

13. G. Mackey, "A remark on locally compact abelian groups," *Bulletin of the American Mathematical Society*, vol. 52 (1946), pp. 940-944.

14. Y. Otobe, "On quasi-evaluations of compact rings," *Proceedings of the Imperial Academy of Tokyo*, vol. 20 (1944), pp. 278-282.

15. Y. Otobe, "Note on locally compact simple rings," *Proceedings of the Imperial Academy of Tokyo*, vol. 20 (1944), p. 200.

16. Y. Otobe, "On locally compact fields," *Japanese Journal of Mathematics*, vol. 19 (1945), pp. 189-202.

17. I. E. Segal, "The group algebra of a locally compact group," *Transactions of the American Mathematical Society*, vol. 61 (1947), pp. 69-105.

18. A. Weil, "L'Intégration dans les Groupes Topologiques et ses Applications," *Actualités Scientifiques et Industrielles*, 869, Paris, 1938.

CORRIGENDA.

ON THE PARTIAL PRODUCTS OF INFINITE PRODUCTS OF ALEPHS.

By F. BAGEMIHL.

p. 207, line 10: the second α should be $\bar{\alpha}$.

p. 207, line 18: $\aleph_\lambda{}^{\overline{\lambda\omega^\delta}}$ should be $\aleph_\lambda{}^{\overline{\omega^\delta}}$.

p. 208, line 19: $\prod_{\zeta<\lambda} (\prod_{\eta<\lambda} \mathfrak{a}_{\tau_{\xi,\eta}})$ should be $\prod_{\xi<\lambda} (\prod_{\eta<\lambda} \mathfrak{a}_{\tau_{\xi,\eta}})$.

p. 210, line 16 should be $\prod_{\xi<\lambda} \mathfrak{p}_\xi = \aleph_{\pi(\omega)}^{\aleph_0} = 2^{\aleph_{\pi(\omega)}} > \mathfrak{p}$.

ADDENDUM TO "A MATRIX DIFFERENTIAL EQUATION OF RICCATI TYPE."*

WILLIAM T. REID

Professor J. Radon has called my attention to the fact that the presentation of the Legendre equation of a problem of Lagrange as a matrix differential equation of Riccati type, and the fundamental theorems on the integration of such an equation, are given in two earlier papers of his ["Über die Oszillationstheoreme der konjugierten Punkte biem Probleme von Lagrange," *Münchner Sitzungsberichte* (1927), pp. 243-257; "Zum Problem von Lagrange," *Abhandlungen aus dem Mathematischen Seminar Hamburg*, vol. 6 (1928), pp. 273-299]. In particular, the results of Sections **3** and **4** of the indicated paper of the author are contained in these papers of Radon.

* This journal, vol. 68 (1946), pp. 237-246.

ON THE ASYMPTOTIC PROBLEMS OF THE ZEROS IN WAVE MECHANICS.*

By Philip Hartman and Aurel Wintner.

1. In the homogeneous, linear differential equations to be considered, the coefficient functions, parameters and integration constants, hence all the solutions, will be assumed to be real-valued, and the trivial solution ($\equiv 0$) will not be called a solution.

In the one-dimensional wave equation

$$\phi'' + \{\lambda - f(s)\}\phi = 0, \tag{1}$$

let $f = f(s)$ be given and continuous for $0 < s < \infty$, and suppose that

$$f(s) \to \infty \qquad \text{as } s \to \infty. \tag{2}$$

Then the coefficient, $\{\ \}$, of ϕ in (1) tends to $-\infty$ for every fixed λ. It follows, therefore, from a theorem of A. Kneser [7] that there belongs to every λ a solution $\phi = \phi_\lambda(s)$ satisfying

$$\phi_\lambda(s) \to 0 \text{ as } s \to \infty, \tag{3}$$

while every solution linearly independent of this particular solution tends to ∞ or $-\infty$. Hence, by Sturm's separation theorem,

$$\phi_\lambda(s) \neq 0 \text{ as } s \to \infty. \tag{4}$$

It also follows that a solution cannot be bounded unless it tends to 0, i. e., that (3) is implied by

$$\limsup_{s\to\infty} |\,\phi_\lambda(s)\,| < \infty. \tag{5}$$

Finally, if two solutions which differ only in a constant factor ($\neq 0$) are considered as identical, the fact mentioned after (3) implies that it is possible to speak of *the* bounded solution of (1) for every fixed λ.

It is easily realized that, since (2) implies that (1) is ultimately "majorized" by the differential equation $\phi'' - C^2\phi = 0$, where C is any given positive constant, the bounded solution of (1) is ultimately majorized by e^{-Cs}. This implies that the bounded solution of (1), besides satisfying (3), must satisfy the (L^2)-condition

* Received December 30, 1947.

$$\int^{\infty} \phi_\lambda^2(s)\,ds < \infty. \tag{6}$$

But (6), where λ is arbitrary, cannot of course be interpreted by saying that "every λ is an eigenvalue," since eigenvalues are selected only by a boundary condition assigned for $s = 0$; a boundary condition which, depending on the behavior of the potential or $f(s)$ near $s = 0$, can be either of a type such as

$$\int_{+0} \phi_\lambda^2(s)\,ds < \infty \text{ or } \limsup_{s\to+0} |\phi_\lambda(s)| < \infty, \tag{6_1}$$

or of the Sturm-Liouville type, e. g.,

$$\phi_\lambda(+0) = 0 \text{ or } \phi'_\lambda(+0) = 0. \tag{6_2}$$

But this issue will not enter into the following considerations, since the results will concern the behavior of the bounded solution of (1) for an arbitrary λ; so that a boundary condition at $s = 0$ need not even be mentioned. If such a boundary condition *is* assigned, the results will be applicable to the particular λ-values which are eigenvalues, but not of course to these λ-values only.

2. This situation is implied by the nature of the results to be proved. In fact, the main result can be formulated as follows:

THEOREM. *For $0 < s < \infty$, let $f(s)$ be a continuous function which tends to ∞ as $s \to \infty$ and has, for large s, a continuous second derivative satisfying*

$$f''(s) \geqq 0 \tag{7}$$

and

$$f''(s) = o\{f'(s)\}^{4/3} \tag{8}$$

(*this o-condition, apparently of a Tauberian nature, actually involves just a certain regularity in the growth of $f'(s)$, as illustrated by the fact that* (8) *is satisfied by every logarithmico-exponential function subject to $f(\infty) = \infty$;* cf. [4], pp. 34-37.)

By virtue of the assumption $f(\infty) = \infty$, the differential equation

$$\phi'' + \{\lambda - f(s)\}\phi = 0$$

has, for every $\lambda = \text{const.}$, one and (to a constant factor) only one solution $\phi = \phi_\lambda(s)$ which tends to 0 as $s \to \infty$. This solution does not vanish as $s \to \infty$ and, if $s = s^*(\lambda)$ is, for large positive λ, the greatest zero of $\phi_\lambda(s)$, the asymptotic form of $s^*(\lambda)$ can, without solving the differential equation, be determined from

(9) $$s^*(\lambda) - f^{-1}(\lambda) \sim \iota/\{3f'(f^{-1}(\lambda))\}^{\frac{1}{3}}, \qquad (\lambda \to \infty),$$

where f^{-1} *denotes the inverse function of the function* f *and* ι *is an absolute constant,*

(10) $$\iota = 1.85575 \cdots.$$

The latter is defined as the least positive root of Airy's case of the cylinder functions, i. e.,

(11) $$\Psi(\iota) = 0 \text{ but } \Psi(t) > 0 \text{ if } 0 \leqq t < \iota,$$

where

(12) $$\Psi(t) = \int_0^\infty \cos(r^3 + rt)\,dr.$$

Needless to say, the existence of the inverse function f^{-1} (for large values of the argument) is assured, since (7) and (2) imply that $f(s)$ is strictly monotone for large s.

The assertion of the Theorem is known for the case of certain particular coefficient functions $f(s)$ in the differential equation (1). For instance, if (1) is a Bessel differential equation, (9) reduces to a result of Watson ([11], p. 521). In addition, Watson's treatment of this particular case, when combined with the complex-analytical method of steepest descent, suggests the asymptotic relation used since Kramers [8] in quantum theory. Along these lines, Zernike [13] has proved (9) for the case in which (1) is Hermite's differential equation.[1] But Watson's device of "trapping" ([11], p. 520), employed by him and Zernike in the case of Bessel and Hermite functions respectively, presupposes a certain amount of explicit estimates which are not available in the general case.

In the proof of the Theorem, this difficulty will be avoided by introducing a new element into the asymptotic consideration, namely, by having recourse to the general existence theorem of A. Kneser [7]. A substantial part of the latter theorem has already been stated above, after (2).

3. A problem related to that of the Theorem was treated by W. E. Milne [9]. His asymptotic formula, replacing (9), supplies the number of zeros,

---

[1] With $\lambda = n$, where n is an integer. But this restriction of λ actually is superfluous in Zernike's calculations, since it is introduced only for the sake of a boundary condition which, in accordance with the remarks made above in connection with (6_0), (6_1), (6_2), has no substantial rôle in the problem. This means that Zernike's result concerning Hermite polynomials can be extended, without any modification of the proof, to the case of the functions of the parabolic cylinder, where λ is not an integer, n.

say $N(\lambda)$, of $\phi_\lambda(s)$, rather than the greatest zero, $s^*(\lambda)$, of $\phi_\lambda(s)$, as $\lambda \to \infty$. Milne's result is contained in, and is substantially equivalent to, the statement that, if $f(s)$ is continuous on the closed half-line $0 \leqq s < \infty$, the relation

$$(13) \qquad \pi N(\lambda) \sim \int^{f^{-1}(\lambda)} \{\lambda - f(s)\}^{\frac{1}{2}} ds, \text{ as } \lambda \to \infty,$$

holds precisely [2] under the assumptions of the above Theorem.

By using the Tauberian restriction (7), it would be possible to deduce (13) from (9), along the lines of the procedure applied by us [6] in the problem of a function N similar to, but somewhat easier than, that of the N which occurs in (13). On the other hand, the converse deduction, that passing from the *accumulated* function $N(\lambda)$ to the *sharp* function $s^*(\lambda)$, is hardly possible.

This suggests that (13) can be deduced under conditions less strict than those needed in the proof of (9). What is less obvious is that the degree of relaxing the restrictions proves to be very considerable. In fact, it can be concluded from the result of [6] that (13) *remains true if only* (2) *and the convexity of* $f(s)$ *are retained, while* (8) *is omitted entirely.* Incidentally, not even the absolute continuity of $f'(s)$, and still less the existence of a continuous $f''(s)$, will be needed.

As all of this has nothing to do with the proof of (9), the proof of (13) under the relaxed conditions will be deferred to the end of the paper.

4. Corresponding to the circumstance that the content of the condition (8) actually is just a condition on the "regular growth" of $f'(s)$, a direct handling of (8) is inconvenient in the proof of the Theorem. More adapted to the use of a "regular growth" is the following set of conditions, which will be referred to as Conditions (*).

CONDITIONS (*). *The function $f(s)$ is continuous for $0 < s < \infty$, tends to ∞ as $s \to \infty$ and has, for large s, a continuous first derivative satisfying*

$$(14) \qquad \iota^3/3 < \mu \leqq \infty,$$

where ι and μ are defined by (10) *and*

[2] Actually, Milne's presentation involves the third, rather than just the second, derivative of $f(s)$, but he points out ([9], footnote on p. 908) that this can be avoided. Such a presentation of Milne's result (11) is that given by Titchmarsh [10], pp. 125-128. Curiously enough, Titchmarsh's presentation, too, assumes not only the existence of a continuous $f''(s)$ but the restriction (8) as well. In the latter regard, cf. the comments in the text below.

(15) $$\mu = \liminf_{s\to\infty} f'(s)s^3,$$

respectively; finally, if $h(s,r)$, *where* r *is independent of* s, *denotes the ratio*

(16) $$h(s,r) = f'(s + r/\{f'(s)\}^{\frac{1}{3}})/f'(s),$$

then

(17) $$h(s,r) \to 1 \text{ as } s \to \infty$$

holds uniformly on every closed bounded r*-interval contained in the portion* $-\mu < r < \infty$ *of the line* $-\infty < r < \infty$.

The possibility of using *these* restrictions on $f(s)$ is due to the fact that, on the one hand,

(i) *Conditions* (*) *are satisfied by every* $f(s)$ *which is continuous for* $0 < s < \infty$ *and has, for large* s, *a continuous second derivative satisfying* (7) *and* (8),

and that, on the other hand,

(ii) *Conditions* (*) *are sufficient for the truth of* (9).

Since the Theorem is a corollary of (i) and (ii) together, only (i) and (ii) will have to be proved.

5. Needless to say, (i) has nothing to do with differential equations.

Proof of (i). According to (2) and (7), $f(s)$ is positive and $f''(s)$ non-negative, hence (for reasons of convexity) $f'(s)$ is positive, if s is large enough, say $s \geqq s_0$. It will be assumed that $s_0 \leqq s < \infty$. Then, since $f'(s)$ is non-decreasing, $f'(s) \geqq f'(s_0) > 0$. Since this implies that $f'(s)s^3 \to \infty$ as $s \to \infty$, it follows that (15) satisfies (14), with $\mu = \infty$. Hence, in order to prove (i), only that part of Conditions (*) remains to be ascertained which concerns (16) and (17).

If a is any positive constant, it is clear from the existence of the above s_0 that (16) defines a function on the half-strip ($s_a \leqq s < \infty$, $-a \leqq r \leqq a$), where s_a is a certain function ($< \infty$) of a. Hence, what remains to be ascertained is that (17) holds uniformly for $-a \leqq r \leqq a$, if $a > 0$ is arbitrarily fixed. But this is sure to be the case if both

(18) $$h(s, -a) \to 1$$

and

(19) $$h(s, a) \to 1$$

hold as $s \to \infty$. In fact, since $f''(t)$ is non-negative, hence $f'(t)$ is non-decreasing, when t is large enough, it is clear from (16) that, if s is large enough, $h(s, r)$ is between $h(s, -a)$ and $h(s, a)$ for every r contained between $r = -a$ and $r = a$. Hence, (18) and (19) suffice in order that (16) be true uniformly for $-a \leqq r \leqq a$.

It is seen from (16) that, if s is large enough, $h(s, -a)$ is positive and that its (real) logarithm is representable in the form

$$\log h(s, -a) = -\int_u^s \{f''(t)/f'(t)\}dt,$$

where $u = u(s)$ is defined by

$$u = s - a/\{f'(s)\}^{\frac{1}{3}}; \qquad u = u(s) < s.$$

Since $u \to \infty$ as $s \to \infty$, and since (8) is assumed, it follows that

$$\log h(s, -a) = o(1) \int_u^s \{f'(t)\}^{\frac{1}{3}}\, dt$$

as $s \to \infty$. But $f'(t)$ is non-decreasing as $t \to \infty$. Consequently, the last integral is majorized by $(s - u)\{f'(s)\}^{\frac{1}{3}}$, which, by the definition of u, is the constant a. Hence, the expression in the last formula line is $o(1)a = o(1)$ as $s \to \infty$. This proves (18).

On the other hand, it is clear from the proof of the last formula line that, as $s \to \infty$,

$$\log h(s, a) = o(1) \int_s^v \{f'(t)\}^{\frac{1}{3}}\, dt,$$

where, corresponding to the definition of u,

$$v = s + a/\{f'(s)\}^{\frac{1}{3}}; \qquad v = v(s) > s.$$

But f' is non-decreasing, and so

$$\int_s^v \{f'(t)\}^{\frac{1}{3}}\, dt \leqq (v - s)\{f'(v)\}^{\frac{1}{3}}.$$

Clearly, the last three formula lines imply that, as $s \to \infty$,

$$\log h(s, a) = o(1)\{f'(v)\}^{\frac{1}{3}}/\{f'(s)\}^{\frac{1}{3}} = o(1)\{h(s, a)\}^{\frac{1}{3}},$$

h being defined by (16).

The last o-estimate of $\log h(s,a)$ can be written in the form

$$\{\log h(s,a)\}/\{h(s,a)\}^{\frac{1}{3}} \to 0 \qquad (s \to \infty).$$

The latter makes it clear that

$$\text{(20)} \qquad \text{either } h(s,a) \to 1 \text{ or } h(s,a) \to \infty.$$

Hence, in order to complete the proof of (19), it is sufficient to rule out the second of the possibilities allowed in (20).

It is sufficient to decide the alternative (20) under the assumption that $f'(s) \to \infty$ as $s \to \infty$. In fact, since $f'(s)$ is ultimately positive and monotone, the negation of $f'(\infty) = \infty$ implies the existence of a finite limit $f'(\infty) > 0$. But (19) is then clear from (16) alone. In the remaining case, where $f'(\infty) = \infty$, the proof proceeds as follows:

If $j(s)$ and $k(s)$ are positive, monotone functions satisfying

$$\int^{\infty} j(s)\,ds < \infty$$

and $k(\infty) = \infty$, respectively, it follows from a classical lemma of Borel, applied by him to the theory of entire functions, that the inequality

$$k(j(k(s)) + s) < 1 + k(s)$$

is satisfied by a "majority" of s-values and so, in particular, at every $s = s_n$ contained in some sequence $s_1, s_2, \cdots$ which tends to ∞. This corollary of Borel's lemma was applied by Biernacki [1] to a question concerning differential equations of the form $x'' + A(s)x = 0$, where $A(\infty) = \infty$. In order to apply it to the present question, choose

$$j(s) = ae^{-s/3} \text{ and } k(s) = \log f'(s).$$

Then the conditions required of $j(s)$ and $k(s)$ are satisfied, $f'(s)$ being monotone and $f'(\infty) = \infty$. Hence, there exists a sequence $s_1, s_2, \cdots$ satisfying $s_n \to \infty$ and

$$\log f'(s_n + a/\{f'(s_n)\}^{\frac{1}{3}}) < 1 + \log f'(s_n).$$

In view of (16), this means that

$$h(s_n, a) < e, \text{ where } n = 1, 2, \cdots \text{ and } s_n \to \infty.$$

Hence, (19) follows from (20).

This completes the proof of (i).

6. The proof of (ii) will be made to depend on facts which can be derived from Kneser's considerations, referred to above. It will be convenient to formulate the facts relevant in this respect in the form of a

LEMMA. *For*

$$0 \leqq t < \infty \tag{21}$$

and for every λ *contained in a range of the form*

$$\lambda_0 \leqq \lambda < \infty, \tag{22}$$

let $F(t,\lambda)$ *be a continuous function of* t *satisfying*

$$F(t,\lambda) \leqq 0 \tag{23}$$

and having the property that the limit function $G(t) = F(t,\infty)$ *exists and the limit relation*

$$F(t,\lambda) \to G(t), \text{ where } \lambda \to \infty, \tag{24}$$

is uniform on every bounded t*-interval of* (21). *Then the differential equation*

$$\ddot{\omega} + F(t,\lambda)\omega = 0 \tag{25}$$

has, for every λ *contained in the range* (22), *a solution* $\omega = \omega_\lambda(t)$ *satisfying the inequalities*

$$\omega_\lambda(t) > 0 \text{ and } \dot{\omega}_\lambda(t) \leqq 0 \text{ for } 0 \leqq t < \infty, \tag{26}$$

and this solution $\omega_\lambda(t)$ *is unique to a positive constant factor, which can be normalized by*

$$\omega_\lambda(0) = 1. \tag{27}$$

In addition, if $\psi(t) = \omega_\infty(t)$ *denotes the corresponding solution of the limiting case,* $\lambda = \infty$, *of* (25), *that is, if* $\psi(t)$ *is that solution of*

$$\ddot{\psi} + G(t)\psi = 0 \tag{28}$$

which satisfies the inequalities

$$\psi(t) > 0 \text{ and } \psi(t) \leqq 0 \text{ for } 0 \leqq t < \infty \tag{29}$$

and is normalized by

$$\psi(0) = 1, \tag{30}$$

(*a solution* ψ *which exists and is unique*), *then, corresponding to* (24), *the limit relation*

$$\omega_\lambda(t) \to \psi(t), \text{ where } \lambda \to \infty, \tag{31}$$

holds uniformly on every bounded t*-interval of* (21).

First, if $f(t)$ is any function which is continuous on (21) and satisfies there the inequality

$$f(t) \leqq 0, \tag{32}$$

then the differential equation

$$\ddot{\phi} + f(t)\phi = 0 \tag{33}$$

has on (21) a *bounded* non-trivial solution, and this solution of (33) is unique if it is normalized by

$$\phi(0) = 1 \tag{34}$$

([7]; cf. also [3], pp. 42-46 and [12], p. 95, and, for a still simpler proof, [5], § 3, footnote). But (33) means that, if $\phi(t)$ is any (real-valued, non-trivial) solution of (33), then the graph of $\phi = \phi(t)$ must always turn its concavities *away* from the t-axis. Hence, if $\phi(t)$ is that solution of (33) which is bounded as $t \to \infty$ and is normalized by (27), then, since its graph runs within the upper half-plane initially, it must always stay there and so, for reasons of convexity, it must satisfy both inequalities

$$\phi(t) > 0, \dot{\phi}(t) \leqq 0 \text{ for } 0 \leqq t < \infty. \tag{35}$$

Next, (35) and (34) become (26) and (27) or (29) and (30) according as (33) is identified with (25) or with (28). These identifications are possible, since, on the one hand, (23) means that (32) is satisfied by $f(t) = F(t, \lambda)$ and, on the other hand, (23) and (24) imply that (32) holds for $f(t) = G(t)$ also.

7. Accordingly, the Lemma will be proved if it is shown that the limit relation (31), along with the uniformity of (31), holds on every bounded t-interval. To this end, recourse will have to be had to the assumption not fully used thus far, namely, to the limit relation (24), along with the uniformity of (24) on every bounded t-interval.

In the proof, two additional facts, listed below as (I) and (II), will be needed for the differential equation (33) in which $f(t)$ is any continuous function satisfying (32) on (21).

(I) If t_1, t_2 is any pair of distinct points of the half-line (21), and if ϕ_1, ϕ_2 is any pair of values, the points (t_1, ϕ_1), (t_2, ϕ_2) of the (t, ϕ)-plane can be joined by exactly one solution arc, $\phi = \phi(t)$, of (33).

(II) If a solution of (33) is defined by the initial conditions $(\phi)_{t=0} = 1$, $(\dot{\phi})_{t=0} = \dot{\phi}_0$, it will have exactly one or no zero on the half-line (21) according

as $-\infty < \dot\phi_0 < \dot\phi(0)$ or $\dot\phi(0) \leqq \dot\phi_0 < \infty$, where $\dot\phi(0)$ denotes the initial slope of the unique solution curve satisfying (34) and (35).

Ad (I). Since (32) is sufficient for the non-existence of "conjugate points," the truth of (I) is contained in the corresponding classical argument concerning Jacobi's equation (cf., e. g., [2], pp. 58-59).

Ad (II). Since (32) implies that every (non-trivial) solution of (32) has at most one zero on the half-line (21), the correctness of the Dedekind cut claimed by (II) can readily be seen from the convexity argument which led to (35) [correspondingly, (II) can be read off from the proof, even though not from the wording, of Kneser's theorem; cf. loc. cit.].

From (I) and (II), it will now be concluded that

$$\liminf_{\lambda\to\infty} \dot\omega_\lambda(0) > -\infty. \tag{36}$$

Although (36) is just the roughest approximation to (31), it turns out to be the crucial step in the proof of (31).

First, (23) and (I) imply that there belongs to every λ on the half-line (22) a unique solution $\omega = \omega(t, \lambda)$ of (25) satisfying

$$\omega(0, \lambda) = 1 \text{ and } \omega(1, \lambda) = 0. \tag{37}$$

Similarly, since (23) and (24) imply that $G(t) \leqq 0$, hence $G(t) - \epsilon < 0$ if $\epsilon > 0$, it follows from (I) that the differential equation

$$\ddot\theta + \{G(t) - \epsilon\}\theta = 0 \tag{38}$$

has a unique solution $\theta = \theta(t; \epsilon)$ satisfying

$$\theta(0, \epsilon) = 1 \text{ and } \theta(1, \epsilon) = 0. \tag{39}$$

Since the coefficient functions in (25) and (38) are non-positive, no non-trivial solution of either of these differential equations can have more than one zero, and so (37) and (38) imply that

$$\omega(t, \lambda) > 0 \text{ and } \theta(t, \epsilon) > 0 \text{ for } 0 < t < 1.$$

Next, if $\epsilon > 0$, then, since (24) is supposed to hold uniformly on every fixed interval $0 \leqq t \leqq$ const., the inequality

$$F(t, \lambda) > G(t) - \epsilon, \text{ where } 0 \leqq t \leqq 1,$$

holds whenever λ is large enough (with reference to ϵ). Clearly, the last two formula lines imply that the integral

$$\int_0^1 \{F(t,\lambda) - G(t) + \epsilon\}\theta(t,\epsilon)\omega(t,\lambda)\,dt$$

is positive. But if (25) and (38) are multiplied by θ and ω, respectively, a subtraction, when followed by a quadrature, shows that this integral is identical with the difference

$$[\dot\theta(t,\epsilon)\omega(t,\lambda) - \dot\omega(t,\lambda)\theta(t,\epsilon)]_0^1.$$

Accordingly, this difference is positive (if λ is large enough with reference to a given $\epsilon > 0$).

Since (37) and (39) reduce this difference to

$$\{0-0\} - \{\dot\theta(0,\epsilon)1 - \dot\omega(0,\lambda)1\} = -\dot\theta(0,\epsilon) + \omega(0,\lambda),$$

it follows that

$$\dot\omega(0,\lambda) > \dot\theta(0,\epsilon).$$

On the other hand, (II) implies that

$$\dot\omega_\lambda(0) > \dot\omega(0,\lambda),$$

since $\omega = \omega(t,\lambda)$ and $\omega = \omega_\lambda(t)$ are two solutions of (25) the first of which has, and the second fails to have, a zero on the half-line (21), as shown by (37) and (26) respectively. But the last two formula lines imply that (if λ is large enough with reference to ϵ)

$$\dot\omega_\lambda(0) > \dot\theta(0,\epsilon),$$

whence the assertion (36) follows (by choosing, for instance, $\epsilon = 1$).

8. It is easy to improve (36) to the assertion that

(40) $$\lim_{\lambda\to\infty} \dot\omega_\lambda(0) \text{ exists}$$

(as a finite limit).

In order to see this, suppose that (40) is false. Then, since (36) and the second of the inequalities (26) imply that $\dot\omega_\lambda(0)$ is bounded as $\lambda \to \infty$, there exist two sequences, say $\lambda_1^*, \lambda_2^*, \cdots$ and $\lambda_1^{**}, \lambda_2^{**}, \cdots$, which tend to ∞ and have the property that, as $n \to \infty$, the value of $\dot\omega_\lambda(0)$ tends to μ^* or to μ^{**} according as $\lambda = \lambda_n^*$ or $\lambda = \lambda_n^{**}$, where μ^* and μ^{**} are two distinct limits. Since the initial condition (27) is independent of λ, the boundedness of the initial condition $\dot\omega_\lambda(0)$ also implies that, if λ_n^* and λ_n^{**} are suitably chosen, the solutions $\omega_\lambda(t)$ belonging to $\lambda = \lambda_n^*$ and $\lambda = \lambda_n^{**}$ will, by equi-

continuity, form sequences which tend, along with their first derivatives, to limit functions, uniformly on every bounded portion of the half-line (21).

Let $\omega^*(t)$ and $\omega^{**}(t)$ denote the respective limit functions. Then, since (24) is supposed to hold uniformly on every bounded portion of the half-line (21), both $\psi=\omega^*(t)$ and $\psi=\omega^{**}(t)$ are solutions of (28). Furthermore, both of these solutions of (28) satisfy (30), by (27), and they are bounded as $t\rightarrow\infty$, by (27) and the second of the inequalities (26), but they are distinct solutions, since $\dot\omega^*(0)=\mu^*$ and $\dot\omega^{**}(0)=\mu^{**}$, where $\mu^*\neq\mu^{**}$. Accordingly, (28) has two distinct *bounded* solutions satisfying (30). But this is impossible, since, as mentioned before (34), a differential equation (33) satisfying (32) cannot have more than one bounded solution normalized by (34). This contradiction proves (40).

It is clear from this proof that $\omega_\lambda(t)$ tends, as $\lambda\rightarrow\infty$, to a limit function, uniformly on every bounded portion of the line (21), and that this limit function is bounded as $t\rightarrow\infty$ and satisfies both (28) and (30). But (28) cannot have more than one bounded solution satisfying (30). Consequently, the limit function must be identical with that solution of (28) which occurs on the right of the limit relation (31), claimed by the Lemma. The proof of the Lemma is therefore complete.

9. It will now be shown that the Lemma readily supplies a

Proof of (ii). Suppose that $f(s)$ in (1) satisfies Conditions (*). Then, since (15), (14) and (10) imply that $f'(s)$ is ultimately positive, and since the same is true, by (2), of $f(s)$ itself, it is possible to choose an s_0 so large that

$$f(s)>0 \text{ and } f'(s)>0 \text{ for } s_0\leqq s<\omega. \tag{41}$$

In particular, if $s=f^{-1}(\lambda)$ denotes the inverse function of the function $\lambda=f(s)$, then $f^{-1}(\lambda)$ exists, is positive and has a positive, continuous first derivative, if $\lambda_0\leqq\lambda<\infty$, where $\lambda_0=f(s_0)$; and

$$f^{-1}(\lambda)\rightarrow\infty \text{ as } \lambda\rightarrow\infty, \tag{42}$$

by (2). It will always be assumed that s, λ exceed s_0, λ_0, respectively.

In particular, (16) defines a positive, continuous function $h(s,r)$ on that portion of the (s,r)-plane which is determined by the inequalities

$$s\geqq s_0,\qquad r\geqq\{f'(s)\}^{\frac{1}{2}}(s_0-s). \tag{43}$$

On the other hand, by (14), there clearly exists a positive number, say δ,

having the property that the inequalities (43) are satisfied whenever s is large enough and

(44) $$-(3^{-\frac{1}{3}}\iota + \delta) \leqq r < \infty,$$

ι being the constant (10). It will be assumed that s_0 is chosen so large as to make (44) and $s \geqq s_0$ together suffice for the second (hence, for both) of the inequalities (43).

For any fixed λ, replace s in (1), and in any solution $\phi = \phi(s)$ of (1), by a new independent variable, $t = as + b$, as follows:

(45) $$t = \{f'(f^{-1}(\lambda))\}^{\frac{1}{3}}\{s - f^{-1}(\lambda)\}, \qquad (s_0 \leqq s < \infty).$$

Then (1) appears in the form (25), where

(46) $$F(t,\lambda) = \{f'(f^{-1}(\lambda))\}^{-\frac{2}{3}}\{\lambda - f(f^{-1}(\lambda) + t/\{f'(f^{-1}(\lambda))\}^{\frac{1}{3}})\}$$

and

(47) $$\omega(t) = \phi(f^{-1}(\lambda) + t/\{f'(f^{-1}(\lambda))\}^{\frac{1}{3}}),$$

$\phi(s)$, $\omega(t)$ being a corresponding pair of solutions of (1), (25).

By the above choice of the bounds λ_0, s_0, the function on the right of (46) is defined for

$$\lambda_0 \leqq \lambda < \infty \text{ and } \{f'(f^{-1}(\lambda))\}^{\frac{1}{3}}\{s_0 - f^{-1}(\lambda)\} \leqq t < \infty$$

(at least). In particular, it is defined whenever λ and $t = r$ satisfy (22) and (44) respectively. Since, if $t = r$, the half-line (44), where $\iota > 0$ and $\delta > 0$, contains the half-line (21), it follows that (46) defines $F(t,\lambda)$ whenever t, λ are on the respective half-lines (21), (22).

10. On the (t,λ)-region (21)-(22), assumption (23) of the Lemma is satisfied. In fact, since (41) implies that the function f, and therefore its inverse function f^{-1}, is positive and increasing, (23) follows from (46).

It will now be verified that the remaining assumption of the Lemma, that concerning (24), is satisfied by $G(t) = -t$. This will prove that the Lemma is applicable, with

(48) $$\ddot{\psi} - t\psi = 0$$

as the limiting case, (28), of (25).

First, since λ is independent of t, differentiation of (46) with respect to t gives

$$\dot{F}(t,\lambda) = -h(f^{-1}(\lambda), t),$$

$h(s,r)$ being defined by (16). But Conditions (*) are assumed, and so their requirement (17) shows that

$$h(f^{-1}(\lambda),t)\rightarrow 1 \text{ as } f^{-1}(\lambda)\rightarrow\infty$$

holds uniformly on every bounded t-interval contained in the half-line (44), where $r=t$. Since the latter half-line contains the half-line (21), where $t=r$, it is now clear from (42) and from the last two formula lines that, as $\lambda\rightarrow\infty$, the limit relation $F(t,\lambda)\rightarrow -1$ holds uniformly on every bounded t-interval of (21). Hence, in order to conclude that the same is true of the limit relation $F(t,\lambda)\rightarrow -t$ (i.e., of (24), where $G(t)=-t$), it is sufficient to apply a quadrature and to observe that the integration constant, $F(0,\lambda)$, of this quadrature vanishes for every λ. In fact, (46) shows that the value of the integration constant is

$$F(0,\lambda)=\{f'(f^{-1}(\lambda))\}^{-\frac{2}{3}}\{\lambda-f(f^{-1}(\lambda))\},$$

and this shows that $F(0,\lambda)\equiv 0$, the second $\{\ \}$ in the last formula being 0 in view of the definition of f^{-1} as the inverse of f.

This proves that the Lemma is applicable when (28) and the coefficient function of (25) are (48) and (46) respectively.

REMARK. Although the wording of the Lemma merely states that (31) holds uniformly on every bounded t-interval contained in the half-line (21), it is clear from the proof of the Lemma that (31) holds uniformly on every bounded t-interval on which (24) is assumed to hold uniformly. Hence, in the present case, (31) holds uniformly on every bounded interval of the half-line (44), where $r=t$. In other words, (31) holds now uniformly not only on every bounded t-interval contained in the half-line (21) but on the (fixed, bounded) t-interval

$$(49)\qquad -(3^{-\frac{1}{3}}\iota+\delta)\leqq t\leqq 0$$

as well.

The proof of the assertion, (9), of (ii) is now straightforward.

First, if $\Psi(t)$ is defined by (12), then, as is well-known (cf. [11], pp. 188),

$$(50)\qquad \psi(t)=\Psi(-3^{\frac{1}{3}}t)/\Psi(0)$$

is that solution of (48) which is bounded as $t\rightarrow\infty$ and is normalized by (30). It follows therefore from the preceding Remark that, by virtue of the linear substitution $t=a_\lambda s+b_\lambda$ defined by (45), the limit relation (31) belonging to (50) holds uniformly on the s-range, the t-image of which is the interval (49). Cf. (47), where the arbitrary solution, $\phi(t)$, of (25) is now chosen to be the particular solution $\phi_\lambda(t)$ specified in the Lemma.

On the other hand, since (11) means that (10) is the least positive zero of (12), the function (50) has a greatest (real) zero, which is $t = -3^{-\frac{1}{3}}\iota < 0$. Since a (real-valued, non-trivial) solution of a (real) linear, homogeneous differential equation of second order can vanish at a point only by changing its sign at that point, and since (31) holds uniformly on the t-interval (49), where $\delta > 0$, it follows that

(I) if λ is large enough, $\omega_\lambda(t)$ has at least one zero on the interval (49), and that

(II) the largest zero of $\omega_\lambda(t)$ on (49) tends to $-3^{-\frac{1}{3}}\iota$ as $\lambda \to \infty$.

In addition, since (31) holds uniformly not only on (49) but on every bounded interval of the half-line (21) as well, it is clear that

(III) if λ is large enough, the largest zero of $\omega_\lambda(t)$ on (49) is the largest real zero of $\omega_\lambda(t)$.

Finally, if $\phi_\lambda(s)$ is defined by placing

(51) $$\phi_\lambda(s) = \omega_\lambda(t) \text{ in virtue of } (45),$$

it is clear, from the definition of $\omega_\lambda(t)$ in the Lemma and from the fact that (1) is identical with (25) in virtue of (45), that $\phi_\lambda(s)$ is a solution of (1) satisfying the requirements specified after (8) and before (9). But $s^*(\lambda)$ in (9) denotes the largest zero of this $\phi_\lambda(s)$. Hence it is seen from (51) and (45) that the three facts, listed before (51) as (I), (II) and (III), imply the truth of (9).

This completes the proof of (ii). Since (i) has been verified before, and since the Theorem is a corollary of (i) and (ii), the Theorem follows.

Appendix.

In this Appendix, (13) will be proved under the assumptions announced after (13) and even under somewhat more general assumptions. This additional generalization, which assumes the continuity of $f(s)$ only on the open half-line, $0 < s < \infty$, is essential from the point of view of the applications. For, on the one hand, the integral (13) is quite sensitive if $f(s)$ is irregular enough as $s \to 0$ and, on the other hand, s usually corresponds to a radius vector, and $f(s)$ is determined by a potential, in the quantum-mechanical problem. Thus, on the one hand, care must be taken of the behavior of $f(s)$ near $s = 0$ and, on the other hand, $f(s)$ cannot reasonably be bounded near $s = 0$. Correspondingly, what will be proved is the following form of

THE "AVERAGED" VARIANT OF THE THEOREM. *On the open half-line*

$$(1)\qquad 0 < s < \infty,$$

let $f(s)$ *be a real-valued, continuous function satisfying*

$$(2)\qquad \liminf f(s) > -\infty \text{ as } s \to 0$$

and

$$(3)\qquad f(s) \to \infty \text{ as } s \to \infty,$$

and suppose that, if s *is large enough, the graph of*

$$(4)\qquad f = f(s) \text{ is convex from below.}$$

Then if $\phi = \phi_\lambda(s)$ *denotes any real-valued solution of*

$$(5)\qquad \phi'' + \{\lambda - f(s)\}\phi = 0,$$

$\phi_\lambda(s)$ *has a finite number of zeros on the half-line* (1) *whenever* λ *is large enough and, if* $N(\lambda)$ *denotes the number of these zeros, the asymptotic form of* $N(\lambda)$ *can a priori be calculated from* f *(and from the inverse function,* f^{-1}*, of* f*), as follows:*

$$(6)\qquad \pi N(\lambda) \sim \Re \int_0^{f^{-1}(\lambda)} \{\lambda - f(r)\}^{\frac{1}{2}} dr \text{ as } \lambda \to \infty.$$

It should be noted that, while (4) (along with either (3) or the continuity of f) implies the existence and the monotony of right- and left-hand derivatives $D^{\pm}f(s)$ everywhere, as well as the identity of the latter nearly everywhere, not even the additional assumption $D^{+}f(s) \equiv D^{-}f(s)$, which is known to be equivalent to the existence of a *continuous* derivative f', suffices for the *absolute* continuity of $D^{\pm}f$. This implies that, although the first derivative of $D^{\pm}f$ or f' exists, by monotony, almost everywhere, f'' is, in general, quite useless in connection with any problem involving asymptotic *integral* estimates. In the sequel, $f'(s)$ will denote either $D^{+}f(s)$ or $D^{-}f(s)$ (if s is large enough, so large that (4) is true).

According to (4) and (3), it is possible to choose an s_0 so large that

$$(7)\qquad f'(s_0) > 0$$

and

$$(8)\qquad df'(s) \geqq 0 \text{ if } s_0 \leqq s < \infty,$$

hence, again by (4),

$$(9)\qquad f'(s) > 0 \text{ if } s_0 \leqq s < \infty.$$

Put $\mu = f(s_0)$. Then $f^{-1}(\lambda)$ exists (as a unique continuous function) for $\mu \leqq \lambda < \infty$. Furthermore,

(10) $$f^{-1}(\lambda) \to \infty \text{ as } \lambda \to \infty,$$

by (3).

It was proved in [6] that if $g(s)$ is a continuous, non-negative function possessing a continuous derivative $g'(s)$ on the closed half-line

(11) $$0 \leqq s < \infty,$$

and if

(12) $$g'(s) = o\{g(s)\}^{\frac{3}{2}} \text{ as } s \to \infty,$$

then

(13) $$\pi n(s) \sim \int_0^s \{g(r)\}^{\frac{1}{2}} dr \text{ as } s \to \infty,$$

where, if $\phi = \phi(s)$ is any (real-valued, non-trivial) solution of the differential equation

(14) $$\phi'' + g(s)\phi = 0,$$

$n(s)$ denotes the number of those zeros of ϕ which are between 0 and s. Since (as observed loc. cit.) the assumption (12) prevents the boundedness of the integral (13) as $s \to \infty$, it is immaterial whether the zeros possibly situated at 0 and/or at s are or are not included in the definition of $n(s)$.

Suppose that (14) is replaced by

(15) $$\phi'' + g(s, \lambda)\phi = 0,$$

where λ is a parameter varying, for instance, over the half-line $\mu < \lambda < \infty$. Let the assumptions of the theorem just quoted be satisfied by $g(s) = g(s, \lambda)$ when λ is fixed, and suppose that assumption (12) is satisfied uniformly for $\mu < \lambda < \infty$, i. e., that

(16) $$\underset{\mu<\lambda<\infty}{\text{l. u. b.}} \mid g'(s, \lambda) \mid / \mid g(s, \lambda) \mid^{\frac{3}{2}} \to 0 \text{ as } s \to \infty$$

(where $f' = \partial f / \partial s$). Then the λ-form of (13) holds uniformly in λ, i. e.,

(17) $$\underset{\mu<\lambda<\infty}{\text{l. u. b.}} \mid 1 - \pi n(s, \lambda) / \int_0^s \{g(r, \lambda)\}^{\frac{1}{2}} dr \mid \to 0 \text{ as } s \to \infty.$$

This is clear from the proof given in [6] for a fixed λ.

On the portion $(0 \leqq s < \infty,\ \mu < \lambda < \infty)$ of the (s, λ)-plane, define a function $g(s, \lambda)$ by placing

$$(18_1) \qquad g(s,\lambda) = \lambda - f(f^{-1}(\lambda) - s) \text{ if } 0 \leqq s \leqq f^{-1}(\lambda) - s_0$$

and

$$(18_2) \qquad g(s,\lambda) = \lambda - f(s_0) \text{ if } f^{-1}(\lambda) - s_0 < s < \infty,$$

where $f(s)$ is the function occurring in (5) and s_0 is the constant defined by (7), (8). It is seen from (18_1) and (18_2) that, while $g(s, \lambda)$ is continuous throughout, it has only right- and left-hand derivatives, $D^{\pm}g(s, \lambda)$, with respect to s. But the latter are monotone throughout and are identical (and continuous) nearly everywhere, and so it is clear from the proof, given loc. cit., that (17) remains true if the uniform o-assumption is meant in the sense that (16) holds when f' denotes $D^{\pm}f$.

It is easy to see that, in this sense, (16) is satisfied for the function $g(s, \lambda)$ defined by (18_1)-(18_2). In fact, from (18_1),

$$(19_1) \qquad g'(s,\lambda) = f'(f^{-1}(\lambda) - s) \text{ if } 0 \leqq s \leqq f^{-1}(\lambda) - s_0$$

and, from (18_2),

$$(19_2) \qquad g'(s,\lambda) = 0 \text{ if } f^{-1}(\lambda) - s_0 < s < \infty;$$

cf. (7), (8) and (9). Since $f(s)$ is absolutely continuous for $s_0 \leqq s < \infty$, it follows from (18_1), (19_1) and (9) that, if $0 < s \leqq f^{-1}(\lambda) - s_0$,

$$g(s,\lambda) = \int_{f^{-1}(\lambda)-s}^{f^{-1}(\lambda)} f'(r)\,dr,$$

hence, by (8),

$$|g(s,\lambda)|^{\frac{3}{2}} \geqq \{f'(f^{-1}(\lambda) - s)\}^{\frac{3}{2}} s^{\frac{3}{2}},$$

and so, by (19_1),

$$|g'(s,\lambda)| / |g(s,\lambda)|^{\frac{3}{2}} \leqq \{f'(f^{-1}(\lambda) - s)\}^{-\frac{1}{2}} s^{-\frac{3}{2}},$$

and therefore, by (7) and (8),

$$(20) \qquad |g'(s,\lambda)| / |g(s,\lambda)|^{\frac{3}{2}} \leqq \{f'(s_0)\}^{-\frac{1}{2}} s^{-\frac{3}{2}}.$$

While this proof of (20) holds only under the proviso, $0 < s \leqq f^{-1}(\lambda) - s_0$, of (19_1), it is seen from (19_2) and (7) that (20) is trivial if this proviso is negated. Hence, (20) holds without any proviso. But (20) implies (16).

Consequently, (17) is applicable. Hence, if $s = s_\lambda$ is any function of λ satisfying

(21) $$s_\lambda \to \infty \text{ as } \lambda \to \infty,$$

then

(22) $$\pi n(s_\lambda, \lambda) \sim \int_0^{s_\lambda} \{g(r,\lambda)\}^{\frac{1}{2}} dr \text{ as } \lambda \to \infty.$$

Choose

(23) $$s_\lambda = f^{-1}(\lambda) - s_0.$$

Then (10) shows that (21) is satisfied.

If (23) and (18_1) are substituted into the integral on the right of (17), it is seen, by introducing the new integration variable $t = f^{-1}(\lambda) - r$, that the integral on the right of (22) is identical with

$$\int_{s_0}^{f^{-1}(\lambda)} \{\lambda - f(t)\}^{\frac{1}{2}} dt.$$

Hence, if t is now denoted by r, and if (23) is substituted on the left of (22), it follows that (22) can be written in the form

(24) $$\pi n(f^{-1}(\lambda) - s_0, \lambda) \sim \int_{s_0}^{f^{-1}(\lambda)} \{\lambda - f(r)\}^{\frac{1}{2}} dr \qquad (\lambda \to \infty).$$

Finally, it is clear from (18_1) that, if $\phi(s)$ is a solution of (15) on the s-interval $0 \leqq s \leqq f^{-1}(\lambda) - s_0$, then $\phi(f^{-1}(\lambda) - s)$ is a solution of (5) for $s_0 \leqq s \leqq f^{-1}(\lambda)$. It follows therefore from the definitions of $N(\lambda)$ and $n(s, \lambda)$ that, with a possible error of ± 1, the difference

(25) $$N(\lambda) - n(f^{-1}(\lambda) - s_0, \lambda)$$

represents the number of zeros of a (real-valued, non-trivial) solution of (5) on the interval $0 < s < s_0$. Since s_0 is a fixed positive constant, it is now seen from the assumption (2) that the difference (25) is asymptotically equal to $1/\pi$ times $\lambda^{\frac{1}{2}} s_0$ and, therefore, to $1/\pi$ times

$$\mathfrak{R} \int_0^{s_0} \{\lambda - f(r)\}^{\frac{1}{2}} dr,$$

as $\lambda \to \infty$. In view of (25), this completes the proof of (6).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Biernacki, "Sur l'équation différentielle $x'' + A(t)x = 0$," *Prace Matematyczno-Fizyczne*, vol. 40 (1933), pp. 163-171.

[2] O. Bolza, *Lectures on the Calculus of Variations*, Chicago, 1904.

[3] E. Borel, *Leçons sur les séries divergentes*, Paris, 1901.

[4] G. H. Hardy, *Orders of Infinity*, Cambridge Tracts, no. 12 (2nd ed., 1924).

[5] P. Hartman, "On differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948).

[6] P. Hartman and A. Wintner, "The asymptotic arcus variation of solutions of linear differential equations of second order," *American Journal of Mathematics*, vol. 70 (1948), pp. 1-10.

[7] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

[8] H. A. Kramers, "Wellenmechanik und halbzahlige Quantisierung," *Zeitschrift für Physik*, vol. 39 (1926), pp. 828-840.

[9] W. E. Milne, "On the degree of convergence of expansions in an infinite interval," *Transactions of the American Mathematical Society*, vol. 31 (1929), pp. 906-918.

[10] E. C. Titchmarsh, *Eigenfunction Expansions*, Oxford, 1946.

[11] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge, 1922.

[12] A. Wintner, "On the Laplace-Fourier transcendents occurring in mathematical physics," *American Journal of Mathematics*, vol. 69 (1947), pp. 87-98.

[13] F. Zernike, "Eine asymptotische Entwicklung für die grösste Nullstelle der Hermiteschen Polynome," *Proceedings of the Royal Academy of Amsterdam*, vol. 34 (1931), pp. 673-680.

LEBESGUE, FRÉCHET AND KERÉKJÁRTÓ VARIETIES.*

By J. W. T. Youngs.[1]

1. Introduction and general theory.

1.1. This paper deals with certain equivalence relations defined over the totality of mappings (= continuous transformations) from a given Peano space (= locally connected continuum) into a fixed metric space. The relations in question are immediate generalizations of concepts employed by Lebesgue, Fréchet and Kerékjártó in the study of surface theory, but before dealing with these specific concepts in detail it is convenient to develop the discussion on a general level.

1.2. For a fixed pair, (X, Y), where X is a Peano space and Y is metric, there is a class, $\mathcal{B}$, consisting of the totality of mappings, f, from X into Y. If E is an equivalence relation over $\mathcal{B}$, then it partitions $\mathcal{B}$ into mutually exclusive equivalence classes known as E-varieties, and the totality of E-varieties forms a class $\mathcal{E}$. It is convenient to make a distinction between an E-variety considered as an element in $\mathcal{E}$ and an E-variety considered as a subset of $\mathcal{B}$. In the former case the symbol $\boldsymbol{E}$ will be employed, in the latter E. There is a natural transformation from $\mathcal{B}$ onto $\mathcal{E}$; namely

$$\phi: \ \mathcal{B} \rightrightarrows \mathcal{E}$$

defined by the requirement that $\phi(f) = \boldsymbol{E}$ if and only if $f \varepsilon E$. (The double arrow indicates that the transformation is onto $\mathcal{E}$.)

1.3. So far nothing has been said about topologies on $\mathcal{B}$ and $\mathcal{E}$. As for $\mathcal{B}$, if ρ is the metric in Y, then it can be metrized by the rule

$$\rho_1\{f_1, f_2\} = \sup_{x \in X} \rho\{f_1(x), f_2(x)\}.$$

(It is to be noted that though $\mathcal{B}$ is now a metric space it is not compact.) In asking for a topology on $\mathcal{E}$ all that is expected is a *limit topology* making $\mathcal{E}$ an L-space. (In other words, the class of convergent sequences and their respective limits is designated and satisfies two requirements: a) If $\boldsymbol{E}_n \rightarrow \boldsymbol{E}_0$ and $\{\boldsymbol{E}_{n_j}\}$ is a subsequence of $\{\boldsymbol{E}_n\}$, then $\boldsymbol{E}_{n_j} \rightarrow \boldsymbol{E}_0$. b) If $\boldsymbol{E}_n = \boldsymbol{E}_0$, $n = 1, 2, 3, \cdots$, then $\boldsymbol{E}_n \rightarrow \boldsymbol{E}_0$.) On the other hand, $\mathcal{E}$ is generated by an

* Received August 2, 1947.

[1] Fellow of the John Simon Guggenheim Memorial Foundation.

equivalence relation over $\mathscr{B}$, consequently it is only natural to demand more than a mere topology on $\mathcal{E}$ in that this topology should have some connection with the topology on $\mathscr{B}$. These demands are that:

1° ϕ be *continuous*; that is, $f_n \to f_0$ imply $\phi(f_n) \to \phi(f_0)$.

2° Each convergent sequence in $\mathcal{E}$ be *covered* by the image of a convergent sequence in $\mathscr{B}$; that is, $\boldsymbol{E}_n \to \boldsymbol{E}_0$ imply that there is a convergent sequence $\{f_n\}$ with limit f_0, and $\phi(f_n) = \boldsymbol{E}_n$, $n = 0, 1, 2, \cdots$.

1.4. There is, of course, no reason why an equivalence relation $\tilde{E}$ over $\mathscr{B}$ should generate a class $\mathcal{E}$ which can be topologized so as to fulfill the above requirements. (Indeed, it will be observed that of the three equivalence relations to be considered in detail, one leads to a class $\mathcal{E}$ which is not subject to such a topology.) In the event the class $\mathcal{E}$ can be topologized in conformity with requirements 1° and 2° of **1.3**, it is said to be properly topologized.

1.5. Two questions are of immediate interest in connection with the general system considered in the above paragraphs:

1° If $\mathcal{E}$ can be properly topologized, is the topology unique?

2° What conditions on $\tilde{E}$ are necessary and sufficient for the existence of a proper topology on $\mathcal{E}$?

1.6. It is quite clear that the answer to the first question is in the affirmative, simply as a consequence of the powerful character of the restrictions in **1.3**. In fact, it will be shown that $\boldsymbol{E}_n \to \boldsymbol{E}_0$ if and only if $0 \neq \underline{\lim}\, E_n \subset E_0$.

First suppose $\boldsymbol{E}_n \to \boldsymbol{E}_0$. Then there is a sequence $f_n \to f_0$ such that $\phi(f_n) = \boldsymbol{E}_n$, $n = 0, 1, 2, \cdots$ by the covering requirement (**1.3**). Consequently, $0 \neq \underline{\lim}\, E_n$. Now if $g_n \to g_0$, $g_n \in E_n$, $n = 1, 2, 3, \cdots$ then $\phi(g_n) \to \phi(g_0)$ by the continuity requirement (**1.3**). But $\phi(g_n) = \boldsymbol{E}_n \to \boldsymbol{E}_0$, therefore $g_0 \in E_0$ and $0 \neq \underline{\lim}\, E_n \subset E_0$.

Conversely, if $0 \neq \underline{\lim}\, E_n \subset E_0$ then there is an $f_n \in E_n$, $n = 0, 1, 2, \cdots$ such that $f_n \to f_0$. Consequently, $\phi(f_n) \to \phi(f_0)$ by the continuity requirement (**1.3**), that is $\boldsymbol{E}_n \to \boldsymbol{E}_0$.

1.7. As to the second question, suppose that the proper topologization (**1.4**) of $\mathcal{E}$ is possible. Then, $f_n \to f_0$, $g_n \to g_0$ and $f_n \tilde{E} g_n$, $n = 1, 2, 3, \cdots$ implies that $\phi(f_n) \to \phi(f_0)$, $\phi(g_n) \to \phi(g_0)$ and $\phi(f_n) = \phi(g_n)$, $n = 1, 2, 3, \cdots$. Consequently, $\phi(f_0) = \phi(g_0)$, that is, $f_0 \tilde{E} g_0$.

Conversely, suppose $\tilde{E}$ has the property that $f_n \to f_0$, $g_n \to g_0$ and $f_n \tilde{E} g_n$, $n = 1, 2, 3, \cdots$ imply $f_0 \tilde{E} g_0$. Define $\boldsymbol{E}_n \to \boldsymbol{E}_0$ if and only if $0 \neq \underline{\lim}\, E_n \subset E_0$.

This rule designates the class of convergent sequences in $\mathcal{E}$; it is necessary to check that the limit of a convergent sequences is unique. If $E_0 \supset \underline{\lim} E_n \neq 0 \neq \underline{\lim} E_n \subset E'_0$, then there is a mapping $f_n \varepsilon E_n$, $n = 1, 2, 3, \cdots$ such that $f_n \rightarrow f_0 \varepsilon E_0$ and a mapping $g_n \varepsilon E_n$, $n = 1, 2, 3, \cdots$ such that $g_n \rightarrow g_0 \varepsilon E'_0$. But $f_n \varepsilon E_n \ni g_n$ implies $f_n \tilde{E} g_n$ and therefore $f_0 \tilde{E} g_0$. Consequently $E_0 = E'_0$ and limits are unique.

Now suppose $\boldsymbol{E}_n \rightarrow \boldsymbol{E}_0$, that is, $0 \neq \underline{\lim} E_n \subset E_0$ and $\{\boldsymbol{E}_{n_j}\}$ is any subsequence of $\{\boldsymbol{E}_n\}$. Then $0 \neq \underline{\lim} E_{n_j} \subset \underline{\lim} E_n \subset E_0$ and $\mathbf{E}_{n_j} \rightarrow \boldsymbol{E}_0$. It is even simpler to see that if $\boldsymbol{E}_n = \boldsymbol{E}_0$, $n = 1, 2, 3, \cdots$, then $\boldsymbol{E}_n \rightarrow \boldsymbol{E}_0$. This shows that the class of convergent sequences and their respective limits designated above makes $\mathcal{E}$ an L-space (**1.** 3). It remains to be shown that the topology is proper (**1.** 4).

If $f_n \rightarrow f_0$, it must be shown that $\phi(f_n) \rightarrow \phi(f_0)$. Suppose that $f_n \varepsilon E_n$, $n = 0, 1, 2, \cdots$, then $f_0 \varepsilon \underline{\lim} E_n$. If $g_n \varepsilon E_n$, $n = 1, 2, 3, \cdots$ and $g_n \rightarrow g_0$, then as $f_n \tilde{E} g_n$, $n = 1, 2, 3, \cdots$ it is true that $f_0 \tilde{E} g_0$ and $g_0 \varepsilon E_0$. Consequently, $0 \neq \underline{\lim} E_n \subset E_0$ and $\boldsymbol{E}_n \rightarrow \boldsymbol{E}_0$. The result follows on observing that $\phi(f_n) = \boldsymbol{E}_n$, $n = 0, 1, 2, \cdots$.

Finally, suppose $\boldsymbol{E}_n \rightarrow \boldsymbol{E}_0$, then $0 \neq \underline{\lim} E_n \subset E_0$ and there must therefore be mappings $f_n \varepsilon E_n$, $n = 0, 1, 2, \cdots$ such that $f_n \rightarrow f_0$. But $\phi(f_n) = \boldsymbol{E}_n$, $n = 0, 1, 2, \cdots$ and so the covering requirement is fulfilled.

1. 8. In accord with the usual terminology, it will be said that the equivalence $\tilde{E}$ generates an *upper semi-continuous decomposition* of $\mathscr{E}$ if and only if $f_n \rightarrow f_0$, $g_n \rightarrow g_0$ and $f_n \tilde{E} g_n$, $n = 1, 2, 3, \cdots$ imply $f_0 \tilde{E} g_0$. The results developed above can now be stated in the following manner: *The collection $\mathcal{E}$ can be properly topologized if and only if the equivalence $\tilde{E}$ generates an upper semi-continuous decomposition of $\mathscr{E}$; moreover the proper topologization of $\mathcal{E}$ is unique.* (These results may be compared with those of Radó-Youngs [9] [2]).

1. 9. This concludes the topological portion of the general theory. A few comments are to be added along what may be called analytic lines. For the contemplated applications of these topological concepts it is convenient to consider a transformation

$$A: \mathscr{E} \rightarrow P$$

where P is the non-negative linear continuum compactified by the addition of a point at infinity. It is not assumed that A is continuous, but rather, in line with the theory of Lebesgue area, that it is *lower semi-continuous*; that is, $f_n \rightarrow f_0$ implies $\liminf A(f_n) \geq A(f_0)$.

---

[2] Numbers in brackets refer to the bibliography.

1.10. A particular pair, $(A, \tilde{E})$ consisting of a lower semi-continuous transformation $A: \mathscr{G} \rightarrow P$ and an equivalence relation $\tilde{E}$ is said to be *coherent* if and only if $f \tilde{E} g$ implies $A(f) = A(g)$. Under these circumstances one can define a transformation $A_E: \mathcal{E} \rightarrow P$ by the rule that $A_E(\boldsymbol{E}) = A(f)$ for any f in E. Moreover, if $\mathcal{E}$ can be properly topologized, then A_E is a lower semi-continuous transformation. For suppose $\boldsymbol{E}_n \rightarrow \boldsymbol{E}_0$, then for each n there is a mapping $f_n \varepsilon E_n$ such that $f_n \rightarrow f_0$. Consequently, $\liminf A_E(\boldsymbol{E}_n) = \liminf A(f_n) \geqq A(f_0) = A_E(\boldsymbol{E}_0)$.

1.11. To make a concluding general remark, if $\tilde{E}$ and $\tilde{F}$ are equivalence relations over $\mathscr{G}$, define $\tilde{E} < \tilde{F}$ to mean $f \tilde{E} g$ implies $f \tilde{F} g$. Now suppose $\boldsymbol{E}$ is any E-variety, then there is an F-variety, $\boldsymbol{F}$, such that $E \subset F$. Consequently the class F may have members, or representations, which are not in E, and as the desirability of having "good" representations needs no apology, the equivalence $\tilde{F}$ may be considered to be superior to the equivalence $\tilde{E}$ simply because F can certainly do no worse (and perhaps do much better) than $\tilde{E}$ at producing equivalence classes with "good" representations. So much for this discussion at the moment; it is time to observe how these general ideas apply to some equivalence relations due to Lebesgue, Fréchet and Kerékjártó.

2. Lebesgue varieties.

2.1. A mapping $f: X \rightarrow Y$ ($=$ from X *into* Y) is said to be Lebesgue *equivalent* to a mapping $g: X \rightarrow Y$ if and only if there is a homeomorphism $h: X \rightrightarrows X$ ($=$ from X *onto* X) such that $f = gh$. (See Lebesgue [5].) The notation to be employed is $f \tilde{L} g$. It is quite obvious that the binary relation L is an equivalence over $\mathscr{G}$ (see 1.2). Suppose that the L-varieties constitute a class $\mathscr{L}$ (see 1.2). The general result of 1.8 will readily check the possibility of a proper topologization (1.4) of $\mathscr{L}$.

2.2. It will be observed that a proper topologization of $\mathscr{L}$ is impossible as Lebesgue equivalence does not generate an upper semi-continuous decomposition of $\mathscr{G}$. To see this suppose X and Y are the same space, namely the closed unit interval, and let

$$f_n(x) = \begin{cases} \dfrac{1}{2^{n-1}}\, x \text{ if } 0 \leqq x \leqq \frac{1}{2} \\ \dfrac{2^n - 1}{2^{n-1}}\,(x-1) + 1 \text{ if } \frac{1}{2} \leqq x \leqq 1 \end{cases}$$

$$f_0(x) = \begin{cases} 0 & \text{if } 0 \leqq x \leqq \frac{1}{2} \\ 2x - 1 & \text{if } \frac{1}{2} \leqq x \leqq 1 \end{cases}$$

$$g_0(x) = x \qquad\qquad 0 \leqq x \leqq 1.$$

It is easy to see that there is a homeomorphism h_n such that $g_0 = f_n h_n$; in fact, since f_n is a homeomorphism, so is f_n^{-1} and $g_0 = f_n f_n^{-1}$, $n = 1, 2, 3, \cdots$. Therefore $f_n \bar{L} g_0$, $n = 1, 2, 3, \cdots$. On the other hand, $f_0 \bar{L} g_0$ is impossible, since f_0 is *not* a homeomorphism while g_0 is. But $f_n \to f_0$ and $g_0, g_0, g_0, \cdots \to g_0$ hence if Lebesgue equivalence were to generate an upper semi-continuous decomposition of $\mathscr{C}$ then $f_0 \bar{L} g_0$ would be mandatory.

This shows that in general the class $\mathscr{L}$ *cannot* have a proper topologization.

2.3. The term equivalence is of course not employed by Lebesgue (nor, for that matter, by Fréchet or Kerékjártó). He states a condition under which two mappings are *not* to be considered as *different*, and this condition is, in the present terminology, the requirement that the two mappings be Lebesgue equivalent (see Lebesgue [5, p. 301]).

On the other hand a little later on, Lebesgue defines convergent sequences and limits in the class $\mathscr{L}$ by a device which may be interpreted to be the following: $\boldsymbol{L}_n \to \boldsymbol{L}_0$ if and only if there is an $f_n \varepsilon L_n$, $n = 0, 1, 2, \cdots$ such that $f_n \to f_0$. Now it is clear that Lebesgue does not wish to permit a convergent sequence $\{\boldsymbol{L}_n\}$ to have two distinct limits, that is, he does not wish to tolerate a situation in which $f_n \bar{L} g_n$, $n = 1, 2, 3, \cdots$, $f_n \to f_0$, $g_n \to g_0$ but $f_0 \bar{L} g_0$ is false. Consequently, one concludes that what Lebesgue ultimately has in mind is a generalization of what has here been defined to be Lebesgue equivalence, a generalization which will guarantee an upper semi-continuous decomposition of $\mathscr{C}$. But for a generalization of Lebesgue equivalence to stand any chance of having this property it must at least be true that if $f_n \bar{L} g_n$, $n = 1, 2, 3, \cdots$, $f_n \to f_0$ and $g_n \to g_0$, then f_0 and g_0 are equivalent in the generalized sense. It is interesting to note that this minimum requirement, this necessary condition, is also sufficient.

2.4. Define $f \bar{M} g$ if and only if there are sequences $\{f_n\}$ and $\{g_n\}$ such that $f_n \to f$, $g_n \to g$ and $f_n \bar{L} g_n$, $n = 1, 2, 3, \cdots$. To show that $\bar{M}$ is an equivalence the transitivity property alone needs to be checked in detail. To this end, suppose $f_1 \bar{M} f_2 \bar{M} f_3$. There are sequences $\{{}_1f_n\}$ and $\{{}_2f_n\}$ such that ${}_1f_n \to f_1$, ${}_2f_n \to f_2$ and ${}_1f_n \bar{L} {}_2f_n$, $n = 1, 2, 3, \cdots$. There are also sequences $\{{}_2\bar{f}_n\}$ and $\{{}_3\bar{f}_n\}$ such that ${}_2\bar{f}_n \to f_2$, ${}_3\bar{f}_n \to f_3$ and ${}_2\bar{f}_n \bar{L} {}_3\bar{f}_n$, $n = 1, 2, 3, \cdots$. The last statement means that for each n there is a homeomorphism $h_n \colon X \rightrightarrows X$ such that ${}_3\bar{f}_n = {}_2\bar{f}_n h_n$. Define ${}_3f_n = {}_2f_n h_n$, $n = 1, 2, 3, \cdots$. Now ${}_2f_n \bar{L} {}_3f_n$ and as ${}_1f_n \bar{L} {}_2f_n$, ${}_1f_n \bar{L} {}_3f_n$ since $\bar{L}$ is transitive. Consequently, $f_1 \bar{M} f_3$ if it can be shown that ${}_3f_n \to f_3$. Notice that

$$\rho_1\{{}_3f_n, f_3\} \leqq \rho_1\{{}_3f_n, {}_3\bar{f}_n\} + \rho_1\{{}_3\bar{f}_n, f_3\}.$$

But

$$\rho_1\{{}_3f_n, {}_3f_n\} = \rho_1\{{}_2f_nh_n, {}_3\bar{f}_n\} = \rho_1\{{}_2f_n, {}_3\bar{f}_nh_n^{-1}\}$$
$$= \rho_1\{{}_2f_n, {}_2\bar{f}_n\} \leq \rho_1\{{}_2f_n, f_2\} + \rho_1\{f_2, {}_2\bar{f}_n\}.$$

(The second inequality is justified in virtue of the fact that if $f: X \to Y$, $g: X \to Y$ and h is a homeomorphism from X onto itself, then

$$\rho_1\{f, g\} = \rho_1\{fh, gh\}$$

because of the definition of the metric on $\mathscr{C}$ (see **1.** 3).) Hence

$$\rho_1\{{}_3f_n, f_3\} \leq \rho_1\{{}_2f_n, f_2\} + \rho_1\{{}_2\bar{f}_n, f_2\} + \rho_1\{{}_3\bar{f}_n, f_3\} \to 0$$

and ${}_3f_n \to f_3$.

It is obvious that $f\,\bar{L}\,g$ implies $f\,\bar{M}\,g$, in other words $L < M$ (see **1.** 11). Now suppose $f_n\,\bar{M}\,g_n$, $n = 1, 2, 3, \cdots$ while $f_n \to f_0$, $g_n \to g_0$. Then for each n there are sequences $\{{}_nf_k\}$ and $\{{}_ng_k\}$ such that ${}_nf_k \to f_n$, ${}_ng_k \to g_n$ and ${}_nf_k\,\bar{L}\,{}_ng_k$, $k = 1, 2, 3, \cdots$. Since $\mathscr{C}$ is metric, for each n one can select a subscript k_n such that $\rho_1\{f_n, {}_nf_{k_n}\} < 1/n > \rho_1\{g_n, {}_ng_{k_n}\}$. Consequently, ${}_nf_{k_n} \to f_0$, ${}_ng_{k_n} \to g_0$ and ${}_nf_{k_n}\,L\,{}_ng_{k_n}$, $n = 1, 2, 3, \cdots$; that is, $f_0\,\bar{M}\,g_0$. Hence $\bar{M}$ generates an upper semi-continuous decomposition of $\mathscr{C}$ and the class, $\mathfrak{M}$, of M-varieties, $\boldsymbol{M}$, can be properly topologized (**1.** 5-**1.** 8).

These comments will be touched upon later when $\bar{M}$ is compared with Fréchet equivalence, the next concept to be defined. It should be stated here, however, that an examination of the viewpoint of Lebesgue has been the subject of several conversations with Radó.

3. Fréchet varieties.

3. 1. A mapping $f: X \to Y$ is Fréchet equivalent to a mapping $g: X \to Y$ if and only if, for every $\epsilon > 0$ there is a homeomorphism $h_\epsilon: X \rightrightarrows X$ such that ρ_1 $f, gh_\epsilon\} < \epsilon$. (See Fréchet [1].) The notation to be employed is $f\,\bar{F}\,g$. It is easy to see that this binary relation is an equivalence over $\mathscr{C}$ and $L < \bar{F}$ (see **1.** 11). The crucial question concerns the possibility of properly topologizing the class, $\mathfrak{F}$, of F-varieties, $\boldsymbol{F}$.

3. 2. THEOREM. *Fréchet equivalence generates an upper semi-continuous decomposition of* $\mathscr{C}$.

Proof. Suppose $f_n\,\bar{F}\,g_n$, $n = 1, 2, 3, \cdots$ while $f_n \to f_0$ and $g_n \to g_0$. For any $\epsilon > 0$ there is a homeomorphism $h_n: X \rightrightarrows X$ such that $\rho_1\{f_n, g_nh_n\} < \epsilon/3$. Now there is an n_0 such that $n > n_0$ implies $\rho_1\{f_n, f_0\} < \epsilon/3 > \rho_1\{g_n, g_0\}$ and in view of the fact that $\rho_1\{g_nh_n, g_0h_n\} = \rho_1\{g_n, g_0\}$ it follows that $\rho_1\{f_0, g_0h_n\} \leq \rho_1\{f_0, f_n\} + \rho_1\{f_n, g_nh_n\} + \rho_1\{g_nh_n, g_0h_n\} < \epsilon$. Consequently $f_0\,\bar{F}\,g_0$.

3. 3. A stronger theorem than the above, however, is possible. An equivalence $\tilde{E}$ over $\mathcal{C}$ is said to generate a *continuous decomposition* of $\mathcal{C}$ if it generates an upper semi-continuous decomposition, and, in addition it is true that $f_n \to f_0 \tilde{E} g_0$ implies there is a sequence $\{g_n\}$ such that $g_n \to g_0$ and $f_n \tilde{E} g_n$, $n = 1, 2, 3, \cdots$.

THEOREM. *Fréchet equivalence generates a continuous decomposition of $\mathcal{C}$.*

Proof. It is only necessary to show that if $f_n \to f_0 \tilde{F} g_0$ then there is a sequence $\{g_n\}$ such that $g_n \to g_0$ and $f_n \tilde{F} g_n$, $n = 1, 2, 3, \cdots$. Since $f_0 \tilde{F} g_0$, for every n there is a homeomorphism $h_n \colon X \rightrightarrows X$ such that $\rho_1\{f_0, g_0 h_n\} < 1/n$. Define $g_n = f_n h_n^{-1}$. Then $g_n \tilde{F} f_n$ and $\rho_1\{g_n, g_0\} = \rho_1\{g_n h_n, g_0 h_n\} = \rho_1\{f_n, g_0 h_n\}$ $\leqq \rho_1\{f_n, f_0\} + \rho\{f_0, g_0 h_n\} < \rho_1\{f_n, f_0\} + 1/n$. Therefore $g_n \to g_0$.

3. 4. The class $\mathfrak{F}$ can therefore be properly topologized by the general methods of the introduction (**1.** 5-**1.** 8). A fact which makes Fréchet equivalence a highly convenient concept, however, is that the proper topologization of $\mathfrak{F}$ metrizes it. This will be shown by making use of the fact that the proper topologization of $\mathfrak{F}$ is unique, (**1.** 6).

3. 5. If $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ are elements of $\mathfrak{F}$ define

$$\rho_2\{\boldsymbol{F}_1, \boldsymbol{F}_2\} = \inf \rho_1\{f_1, f_2\}$$

where the infimum is taken over mappings f_1 and f_2 such that $f_1 \varepsilon F_1$ while $f_2 \varepsilon F_2$. To show that ρ_2 is a metric on $\mathfrak{F}$ it is only necessary to check the triangle inequality and the assertion that $\rho_2\{\boldsymbol{F}_1, \boldsymbol{F}_2\} = 0$ implies $\boldsymbol{F}_1 = \boldsymbol{F}_2$.

Concerning this last statement, suppose $\rho_2\{\boldsymbol{F}_1, \boldsymbol{F}_2\} = 0$. Select $\bar{f}_i \varepsilon F_i$, $i = 1, 2$. For any $\epsilon > 0$, there is a mapping $f_i \varepsilon F_i$, $i = 1, 2$, such that $\rho_1\{f_1, f_2\} < \epsilon/3$. From the definition of Fréchet equivalence there is homeomorphism $h_i \colon X \rightrightarrows X$ such that $\rho_1\{f_i, \bar{f}_i h_i\} < \epsilon/3$, $i = 1, 2$. Consequently, $\rho_1\{\bar{f}_1 h_1, \bar{f}_2 h_2\} \leqq \rho_1\{\bar{f}_1 h_1, f\ \} + \rho_1\{f_1, f_2\} + \rho_1\{f_2, f_2 h_2\} < \epsilon$. Hence $\rho_1\{\bar{f}_1, \bar{f}_2 h_2 h_1^{-1}\}$ $< \epsilon$ and $\bar{f}_1 \tilde{F} f_2$. This shows that $\boldsymbol{F}_1 = \boldsymbol{F}_2$.

Relative to the triangle inequality, suppose $\boldsymbol{F}_1, \boldsymbol{F}_2$ and $\boldsymbol{F}_3$ are elements of $\mathfrak{F}$. Given $\epsilon > 0$ there is a mapping $f_i \varepsilon F_i$, $i = 1, 2$, such that $\rho_2\{\boldsymbol{F}_1, \boldsymbol{F}_2\}$ $> \rho_1\{f_1, f_2\} - \epsilon/3$ and a mapping $\bar{f}_i \varepsilon F_i$, $i = 2, 3$ such that $\rho_2\{\boldsymbol{F}_2, \boldsymbol{F}_3\}$ $> \rho_1\{\bar{f}_2, \bar{f}_3\} - \epsilon/3$. But there is a homeomorphism $h \colon X \rightrightarrows X$ such that $\rho_1\{\bar{f}_2, f_2 h\} < \epsilon/3$. Now $\rho_2\{\boldsymbol{F}_1, \boldsymbol{F}_2\} + \rho_2\{\boldsymbol{F}_2, \boldsymbol{F}_3\} > \rho_1\{f_1, f_2\} + \rho_1\{\bar{f}_2, \bar{f}_3\} - 2\epsilon/3$ $> \rho_1\{f_1 h, f_2 h\} + \rho_1\{f_2 h, \bar{f}_2\} + \rho_1\{\bar{f}_2, \bar{f}_3\} - \epsilon \geqq \rho_1\{f_1 h, \bar{f}_3\} - \epsilon \geqq \rho_2\{\boldsymbol{F}_1, \boldsymbol{F}_3\} - \epsilon$. Consequently the triangle inequality is satisfied.

It remains to be shown that this metric on $\mathfrak{F}$ is a proper topology (**1.** 4). Suppose $f_n \to f_0$, then $\rho_2\{\phi(f_n), \phi(f_0)\} \leqq \rho_1\{f_n, f_0\}$ since $\boldsymbol{F}_n = \phi(f_n)$ implies

$f_n \varepsilon F_n$ (see **1.2**). Hence $\phi(f_n) \to \phi(f_0)$ and ϕ is continuous. Next, suppose that $\boldsymbol{F}_n \to \boldsymbol{F}_0$, that is $\rho_2\{\boldsymbol{F}_n, \boldsymbol{F}_0\} \to 0$. Let $f_0 \varepsilon F_0$, then for each n there is an $f_n \varepsilon F_n$ and ${}_0f_n \varepsilon F_0$ such that $\rho_1\{\bar{f}_n, \bar{f}_n\} \leq 2\rho_2\{\boldsymbol{F}_n, \boldsymbol{F}_0\}$. Moreover, there is a homeomorphism h_n: $X \rightrightarrows X$ such that $\rho_1\{f_0, {}_0f_nh_n\} < 1/n$, $n = 1, 2, 3, \cdots$. Consequently, $\rho_1\{\bar{f}_nh_n, f_0\} \leq \rho_1\{\bar{f}_nh_n, {}_0f_nh_n\} + \rho_1\{{}_0f_nh_n, f_0\} < 2\rho_2\{\boldsymbol{F}_n, \boldsymbol{F}_0\} + 1/n$. If $f_n = \bar{f}_nh_n$, then $f_n \varepsilon F_n$, $n = 1, 2, 3, \cdots$, and $f_n \to f_0$. Hence the covering requirement of **1.3** is fulfilled and the metric topology of $\mathfrak{F}$ is proper.

3.6. In view of the conclusion above and the fact that there is precisely one proper topology on $\mathfrak{F}$ the general method of topologization in **1.6** is the topology of the metric ρ_2. This metric is known, by the way, as the Fréchet metric and is defined essentially in the same manner as in Fréchet's paper [1].

3.7. Of interest in comparing the results of Lebesgue with those of Fréchet is the following

THEOREM. *Fréchet equivalence is the same as the generalized Lebesgue equivalence of* **2.4**.

Proof. If $f \tilde{F} g$ then there is a homeomorphism h_n: $X \rightrightarrows X$ such that $\rho_1\{f, gh_n\} < 1/n$, $n = 1, 2, 3, \cdots$. That is, $gh_n \to f$. But, by definition, $g \tilde{L} gh_n$ while $g, g, g, \cdots \to g$. Hence $f \tilde{M} g$.

If $f \tilde{M} g$, then there are sequences $\{f_n\}$ and $\{g_n\}$ such that $f_n L g_n$, $n = 1, 2, 3, \cdots$, and $f_n \to f$, $g_n \to g$. Hence there is a homeomorphism h_n: $X \rightrightarrows X$ such that $f_n = g_nh_n$, $n = 1, 2, 3, \cdots$. Now $\rho_1\{f, gh_n\} \leq \rho_1\{f, f_n\} + \rho_1\{f_n, g_nh_n\} + \rho_1\{g_nh_n, gh_n\} = \rho_1\{f, f_n\} + \rho_1\{g_n, g\} \to 0$. This guarantees that $f \tilde{F} g$.

3.8. It is of historical interest to remark that (with a suitable interpretation of some of his remarks) it is apparent that Lebesgue eventually applies the equivalence $\tilde{M}$ to the analytic problems he has in mind and topologizes $\mathfrak{M}$ in the manner of **1.6**. On the other hand, Fréchet works with $\tilde{F}$ and metrizes $\mathfrak{F}$ essentially in the manner of **3.5**. The preceding paragraphs have shown that $\mathfrak{M} \equiv \mathfrak{F}$ when considered both as partitions of $\mathfrak{C}$ and as spaces. There appears to be a real advantage, however, to a direct definition of a metric on $\mathfrak{F}$ rather than a topologization in the manner of **1.6**, in spite of the fact that from the topological point of view they are the same. Certainly, in the history of the subject of area, salient advances were made after Fréchet defined the metric which now bears his name, and though it would no doubt be an error to attribute this to the neater form of a topology by means of a metric, it appears safe to consider that this feature

did play some part in the rapid subsequent development of the theory of surface area.

4. Kerékjártó varieties.

4.1. The definition of Kerékjártó equivalence requires a theorem on factorization of a mapping. (See Whyburn [11]) If $f: X \to Y$ is a mapping, then there is a Peano space $\mathfrak{M}$, a montone mapping $m: X \overset{\to}{\to} \mathfrak{M}$ and a light mapping $l: \mathfrak{M} \to Y$ such that $f = lm$. The composition, or product, lm, is said to be a monotone-light factorization of f with middle space $\mathfrak{M}$. If $l_1 m_1$ is a monotone light factorization of f with middle space $\mathfrak{M}_1$, then there is a unique homeomorphism $h: \mathfrak{M} \rightleftarrows \mathfrak{M}_1$ such that $hm = m_1$ and $l = l_1 h$.

4.2. Two mappings $f_1: X \to Y$ and $f_2: X \to Y$ are Kerékjártó equivalent if and only if for *each* monotone-light factorization $l_i m_i$ of f_i with middle space $\mathfrak{M}_i$, $i = 1, 2$, there is a homeomorphism $h: \mathfrak{M}_1 \rightleftarrows \mathfrak{M}_2$ such that $l_1 = l_2 h$. (See Kerékjártó [4].) The notation to be employed is $f_1 \bar{K} f_2$. It is not difficult to see that if $l_i m_i$ is some *particular* monotone-light factorization of f_i with middle space $\mathfrak{M}_i$, $i = 1, 2$, and there is a homeomorphism $h: \mathfrak{M}_1 \rightleftarrows \mathfrak{M}_2$ with the property that $l_1 = l_2 h$, then $f_1 \bar{K} f_2$. Thus it is a simple matter to prove that $f_1 \bar{K} f_2$ if and only if there is a monotone-light factorization lm_i of f_i with middle space $\mathfrak{M}$, $i = 1, 2$. (For details one should consult Radó [7]).

4.3. There is no strain in showing that the binary relation $\bar{K}$ justifies the terminology employed on its behalf and is, indeed, an equivalence; it is not so easy to prove that the resulting decomposition of $\mathscr{C}$ is upper semicontinuous. Before passing on to the proof of this statement it is necessary to consider a few preliminary ideas stated in the following paragraphs.

4.4. If: 1° $g_n: X \to Y$ is a mapping, $n = 0, 1, 2, \cdots$; 2° $g_n \to g_0$; 3° $X_n \subset X$, $n = 0, 1, 2, \cdots$; 4° $\lim X_n = X_0$; then $\lim g_n(X_n) = g_0(X_0)$. (The space X need only be a compactum.)

4.5. If: 1° $m_n: X \to Y$ is a mapping, $n = 0, 1, 2, \cdots$; 2° m_n is monotone for $n > 0$; 3° $m_n \to m_0$; 4° $m_n(X) = \mathfrak{M}_n$, $n = 0, 1, 2, \cdots$; then m_0 is monotone if and only if $\mathfrak{M}_n$ converges to $\mathfrak{M}_0$ 0-regularly (Whyburn [10]).

4.6. Suppose that Z is the space consisting of the real numbers $0, 1, \frac{1}{2}, \cdots$ and for future convenience, let $z_0 = 0$, $z_n = 1/n$, $n = 1, 2, 3, \cdots$. Then the product space $(X \times Z)$ is a compactum, it has a denumerable number of components $(X \times z_n)$, $n = 0, 1, 2, \cdots$, and each component is Peanian.

4.7. If $\{f_n\}$ is a sequence of transformations $f_n\colon X \to Y$, $n = 0, 1, 2, \cdots$, then it generates a transformation $F\colon (X \times Z) \to Y$ defined as follows: for $x \varepsilon X$, $F(x, z_n) = f_n(x)$, $n = 0, 1, 2, \cdots$. (Nothing is assumed about the continuity of these transformations; when continuous they are called mappings.) Conversely, if $F\colon (X \times Z) \to Y$ is a transformation, it generates a sequence $\{f_n\}$ of transformations $f_n\colon X \to Y$, $n = 0, 1, 2, \cdots$ defined in the following manner: for $x \varepsilon X$, $f_n(x) = F(x, z_n)$, $n = 0, 1, 2, \cdots$.

4.8. If $f_n \to f_0$ and all the transformations are mappings, then F is likewise a mapping.

4.9. If F is a mapping, then $f_n \to f_0$ and all these transformations are mappings.

4.10. Now suppose that f_n is a sequence of mappings $f_n\colon X \to Y$, $n = 0, 1, 2, \cdots$ such that $f_n \to f_0$. Then the generated mapping $F\colon (X \times Z) \to Y$ has a monotone-light factorization LM with middle space $\mathfrak{M}$. (The space X need only be a compactum for the truth of the statements in **4.1**.) Moreover, it is easy to see that $\mathfrak{M}$ is a compactum consisting of a sequence $\{\mathfrak{M}_n\}$ of components where $\mathfrak{M}_n$ is the image of $(X \times z_n)$ under the mapping M and is consequently Peanian, $n = 0, 1, 2, \cdots$.

4.11. Consider the sequence $\{m_n\}$ of mappings $m_n\colon X \to \mathfrak{M}$, $n = 0, 1, 2, \cdots$ generated by the mapping $M\colon (X \times Z) \rightrightarrows \mathfrak{M}$. It follows that:

1° $m_n\colon X \rightrightarrows \mathfrak{M}_n$, $n = 0, 1, 2, \cdots$ (see **4.7** and **4.10**.)

2° m_n is monotone, $n = 0, 1, 2$ (see **4.9** and **4.10**.)

3° $m_n \to m_0$ (see **4.9**.)

4° $\lim \mathfrak{M}_n = \mathfrak{M}_0$ and the convergence is 0-regular (see **4.5**.).

4.12 Suppose $l_n\colon \mathfrak{M}_n \to Y$ is the mapping $L\colon \mathfrak{M} \to Y$ restricted to $\mathfrak{M}_n$, $n = 0, 1, 2, \cdots$. It follows that

1° If $g_n\colon X \rightrightarrows \mathfrak{M}_n$ is monotone, $n = 1, 2, 3, \cdots$, $g_0\colon X \to \mathfrak{M}$ is a mapping and $g_n \to g_0$, then $g_0\colon X \rightrightarrows \mathfrak{M}_0$, g_0 is monotone and $l_n g_n \to l_0 g_0$. (By **4.4**, $g_0(X) = \lim g_n(X)$ and by **4.11** this is $\lim \mathfrak{M}_n = \mathfrak{M}_0$. Since the convergence is 0-regular the mapping g_0 is monotone by **4.5**. The transformation $G\colon (X \times Z) \to \mathfrak{M}$ generated by the sequence $\{g_n\}$ is continuous by **4.8**, hence $LG\colon (X \times Z) \to Y$ is continuous. But $\{l_n g_n\}$ is the sequence generated by LG. Hence $l_n g_n \to l_0 g_0$ by **4.9**.)

2° If $\mathfrak{X}_n \subset \mathfrak{M}_n$, $n = 0, 1, 2, \cdots$ and $\lim \mathfrak{X}_n = \mathfrak{X}_0$ then $\lim l_n(\mathfrak{X}_n) = l_0(\mathfrak{X}_0)$. (Notice that $L, L, L, \cdots \to L$ and apply **4.4**.)

3° l_n is light, $n = 0, 1, 2, \cdots$ (L is light.)

4° $f_n = l_n m_n$, $n = 0, 1, 2, \cdots$ $(F = LM.)$

4.13. With this introduction the principal theorem of this section can now be proved.

THEOREM. *Kerékjártó equivalence generates an upper semi-continuous decomposition of $\mathfrak{F}$.*

Proof. Suppose $f_n{}^i \to f_0{}^i$, $i = 1, 2$, and $f_n{}^1 \bar{K} f_n{}^2$, $n = 1, 2, 3, \cdots$. It will be shown that $f_0{}^1 \bar{K} f_0{}^2$.

4.14. In a manner which has been indicated in the preceding paragraphs, the sequence $\{f_n{}^i\}$ gives rise to the sequence $\{m_n{}^i\}$ of monotone mappings $m_n{}^i\colon X \overrightarrow{\rightarrow} \mathfrak{M}_n{}^i$ and the sequence $l_n{}^i$ of light mappings $l_n{}^i\colon \mathfrak{M}_n{}^i \to Y$ connected by the formula $f^i = l_n{}^i m_n{}^i$, $n = 0, 1, 2, \cdots$, $i = 1, 2$. (It is recalled that $\mathfrak{M}_0{}^i, \mathfrak{M}_1{}^i, \mathfrak{M}_2{}^i, \cdots$ are the components of a compactum $\mathfrak{M}^i$, $i = 1, 2$.)

Since $f_n{}^1 \bar{K} f_n{}^2$ there is a homeomorphism $h_n\colon \mathfrak{M}_n{}^1 \overrightarrow{\rightarrow} \mathfrak{M}_n{}^2$ such that $l_n{}^1 = l_n{}^2 h_n$, $n = 1, 2, 3, \cdots$.

4.15. *Assertion.* The sequence $\{h_n m_n{}^1\}$ is equicontinuous.

Remark. The above statement has the flavor of a fundamental lemma of Radó [7, p. 425]. He shows that if a sequence $\{lm_n\}$ is equicontinuous, where $m_n\colon X \overrightarrow{\rightarrow} \mathfrak{X}$, $n = 1, 2, 3, \cdots$ and $l\colon \mathfrak{X} \to Y$ is light, then $\{m_n\}$ is equicontinuous.

4.16. If the assertion is false, then it may be assumed that there is an $\epsilon > 0$ and a couple of sequences $\{x_n\}$ and $\{y_n\}$ such that $x_n \to x_0 \leftarrow y_n$ and $d\{h_n m_n{}^1(x_n \cup y_n)\} > \epsilon$, $n = 1, 2, 3, \cdots$. But $m_n{}^1 \to m_0{}^1$ by **4.11**, 3°, hence $\lim m_n{}^1(x_n \cup y_n) = m_0{}^1(x_0)$ by **4.4**. Consequently, $d\{m_n{}^1(x_n \cup y_n)\} \to 0$. Moreover, $\lim \mathfrak{M}_n{}^1 = \mathfrak{M}_0{}^1$ and the convergence is 0-regular (see **4.11**, 4°). Hence for each $n = 1, 2, 3, \cdots$ there is a continuum $\mathfrak{C}_n{}^1$ such that $\mathfrak{M}_n{}^1 \supset \mathfrak{C}_n{}^1 \supset m_n{}^1(x_n \cup y_n)$, and $d(\mathfrak{C}_n{}^1) \to 0$. Consequently

$$\lim \mathfrak{C}_n{}^1 = m_0{}^1(x_0).$$

Notice that $\mathfrak{C}_n{}^2 \equiv h_n(\mathfrak{C}_n{}^1) \supset h_n m_n{}^1(x_n \cup y_n)$ and so $d(\mathfrak{C}_n{}^2) > \epsilon$, $n = 1, 2, 3, \cdots$. Select a convergent subsequence $\{\mathfrak{C}_{n_j}{}^2\}$ from the sequence $\{\mathfrak{C}_n{}^2\}$. If $\lim \mathfrak{C}_{n_j}{}^2 = \mathfrak{C}_0{}^2$, then $d(\mathfrak{C}_0{}^2) \geqq \epsilon$, and $\mathfrak{C}_0{}^2$ is a continuum. Now $\lim l_{n_j}{}^2(\mathfrak{C}_{n_j}{}^2) = l_0{}^2(\mathfrak{C}_0{}^2)$ by **4.12**, 2°, but $l_{n_j}{}^2(\mathfrak{C}_{n_j}{}^2) = l_{n_j}{}^2 h_{n_j}(\mathfrak{C}_{n_j}{}^1) = l_{n_j}{}^1(\mathfrak{C}_{n_j}{}^1)$, $j = 1, 2, 3, \cdots$ by **4.14**. On the other hand, $\lim l_{n_j}{}^1(\mathfrak{C}_{n_j}{}^1) = l_0{}^1(\lim \mathfrak{C}_{n_j}{}^1)$ by **4.12**, 2° and $l_0{}^1(\lim \mathfrak{C}_{n_j}{}^1) = l_0{}^1 m_0{}^1(x_0) = f_0{}^1(x_0)$. Hence

$$l_0{}^2(\mathfrak{C}_0{}^2) = f_0{}^1(x_0).$$

But this is in contradiction to the lightness of $l_0{}^2$ since $\mathfrak{C}_0{}^2$ is a non-degenerate continuum.

4.17. To complete the proof of the theorem, since $\{h_n m_n^1\}$ is an equicontinuous sequence it is a classical result that it contains a convergent subsequence in $\mathscr{C}$, see **1.2**. For the purpose in mind, there is no loss of generality in assuming that the sequence itself is convergent, that is,

$$g_n \equiv h_n m_n^1 \to g_0$$

In virtue of **4.12**, 1°, $g_0\colon X \rightrightarrows \mathfrak{M}_0$, g_0 is monotone and $l_n^2 g_n \to l_0^2 g_0$. But $l_n^2 g_n = l_n^2 h_n m_n^1 = l_n^1 m_n = f_n^1 \to f_0^1$. Therefore $f_0^1 = l_0^2 g_0$, and as $f_0^2 = l_0^2 m_0^2$ the mappings f_0^1 and f_0^2 are Kerékjártó equivalent by **4.2**.

4.18. Now as it is certain that the equivalence $\bar{K}$ generates an upper semi-continuous decomposition of $\mathscr{C}$, the general theory of **1.5-1.8** shows that there is a proper topologization of the class $\mathscr{K}$ of K-varieties $\boldsymbol{K}$. It is not known whether the proper topologization of $\mathscr{K}$ yields a metric space or not.

4.19. It will be recalled that Fréchet equivalence generated a continuous decomposition of $\mathscr{C}$, and it is interesting to compare this state of affairs with the situation which occurs under Kerékjártó equivalence. It will be shown that Kerékjártó equivalence does not generate a continuous decomposition of $\mathscr{C}$. Consider a polar coordinate system (r, θ) in Euclidean 3-space. The space X is the totality of points (r, θ) such that $r \leq 1$, while the space Y is the totality of points (r, θ). Consider mappings $f_n\colon X \to Y$, $n = 1, 2, 3, \cdots$ defined as follows:

$$f_n(r, \theta) = \begin{cases} (-r \sin \theta + r/n, \theta), & -\pi \leq \theta \leq 0, \quad 0 \leq r \leq 1. \\ (r \sin \theta + r/n, \theta), & 0 \leq \theta \leq \pi, \quad 0 \leq r \leq 1. \end{cases}$$

and a mapping $f_0\colon X \to Y$ given by the formula

$$f_0(r, \theta) = \begin{cases} (-r \sin \theta, \theta), & -\pi \leq \theta \leq 0, \quad 0 \leq r \leq 1. \\ (r \sin \theta, \theta), & 0 \leq \theta \leq \pi, \quad 0 \leq r \leq 1. \end{cases}$$

It is clear that $f_n \to f_0$. Now consider a mapping $g_0\colon X \to Y$ defined as follows

$$g_0(r, \theta) = \begin{cases} (-r \sin \theta, \theta), & -\pi \leq \theta \leq 0, \quad 0 \leq r \leq 1. \\ (r \sin \theta, \pi - \theta), & 0 \leq \theta \leq \pi, \quad 0 \leq r \leq 1. \end{cases}$$

It is obvious that $f_0 \bar{K} g_0$ since each of these mappings is monotone and $f_0(X) = g_0(X)$, (**4.2**). On the other hand, suppose that there is a sequence $\{g_n\}$ of mappings $g_n\colon X \to Y$ such that $g_n \bar{K} f_n$, $n = 1, 2, 3, \cdots$ and $g_n \to g_0$. For convenience, let $\dot{g}_n$ be the mapping g_n restricted to the circumference, $\dot{X}$, of the circle X, $n = 0, 1, 2, \cdots$. The notation $\dot{f}_n$ has an analogous meaning. It is important to notice that, since $g_n \bar{K} f_n$ and f_n is a homeomorphism from the closed 2-cell X onto a closed 2-cell $f_n(X)$, g_n must be a monotone mapping

onto the closed 2-cell $f_n(X)$, $n = 1, 2, 3, \cdots$. Consequently $\dot{g}_n$ is monotone and $\dot{g}_n(\dot{X}) = \dot{f}_n(\dot{X})$, $n = 1, 2, 3, \cdots$ (see Whyburn [11, pp. 165 and 173]). Now suppose that X and Y are given the standard counter-clockwise orientations, and consider the mapping g_n, $n = 1, 2, 3, \cdots$. It follows that the index of $\dot{g}_n$ is the same integer $k \neq 0$ at each point of the bounded component of $Y - \dot{g}_n(\dot{X})$. But $g_n \to g_0$, hence $\dot{g}_n \to \dot{g}_0$, a situation which implies that the index of $\dot{g}_0$ is also k at each point of the two bounded components of $Y - \dot{g}_0(\dot{X})$. On the other hand, the definition of g_0 shows that $\dot{g}_0$ has an index 1 at each point of one of the components in question, and -1 in the other. This contradiction shows that Kerékjártó equivalence does not generate a continuous decomposition of $\mathcal{B}$.

4.20. This section is concluded with the statement of a theorem which is of considerable historical interest and will be employed in the next section.

THEOREM (Kerékjártó). *If $f \tilde{F} g$ then $f \tilde{K} g$.*

Remark. Kerékjártó [4] proved this in case X is a 2-sphere. His method was employed by Youngs [13] in generalizing the result to the case in question. Radó [7] gives the most satisfactory proof using a key lemma which has already been mentioned in **4.15**. Kerékjártó was under the impression that the converse is true if X is a 2-sphere; in fact it is false, see Youngs [13]. In this condition it should be mentioned, however, that if $f \tilde{K} g$, where both f and g are mappings from a 2-sphere, and, in addition, satisfy a certain homology condition, then $f F g$.

5. Surfaces and Lebesgue area.

5.1. The discussion in this section is concerned with a *special* class, $\mathcal{B}$, of mappings, $f\colon X \to Y$, and a lower semi-continuous transformation $A\colon \mathcal{B} \to P$ (see **1.9**). The particular transformation to be considered has the property that each one of the three pairs $(A, \tilde{L})$, $(A, \tilde{F})$ and $(A, \tilde{K})$ is coherent (**1.10**). Consequently, transformations $A_L\colon \mathcal{L} \to P$, $A_F\colon \mathcal{F} \to P$ and $A_K\colon \mathcal{K} \to P$ can be defined as in **1.10**. In view of the fact that $\mathcal{F}$ and $\mathcal{K}$ can be properly topologized (**3.4** and **4.18**) the transformations A_F and A_K are lower semi-continuous (**1.10**). Thus it would appear that $\mathcal{F}$ and $\mathcal{K}$ display certain advantages not enjoyed by $\mathcal{L}$. Moreover, Lebesgue equivalence implies Fréchet equivalence (**3.1**), which in turn implies Kerékjártó equivalence (**4.20**); hence, in the terminology of **1.11**, $\tilde{L} < \tilde{F} < \tilde{K}$. In other words there is evidence to indicate that Kerékjártó equivalence is "superior" (**1.11**) both to Fréchet and Lebesgue equivalence. On the other hand it will be seen that from the point of view of applications there are strong reasons for preferring Fréchet equivalence.

5.2. The first point of specialization to consider is the class $\mathscr{C}$. To this end, the pair (X, Y) is considerably restricted; in fact X is understood to be a 2-cell (or 2-sphere) while Y is Euclidean 3-space. The restriction on Y to Euclidean 3-space rather than Euclidean n-space is not essential; the restriction on X is of definite importance. With this understanding as to the pair (X, Y), an L-variety will be called a Lebesgue surface, an F-variety, a Fréchet surface, and a K-variety, a Kerékjártó surface.

5.3. The special transformation $A\colon \mathscr{C} \to P$ now to be considered will lead to the Lebesgue area of Fréchet surfaces. The definition of A has been covered in some detail elsewhere, see Radó [6] or Youngs [12 and 13] and a précis of the ideas should suffice here. A special class, $\mathfrak{P}$, of mappings, $p\colon X \to Y$, known as polyhedral mappings is defined. There is no general agreement in the literature as to the precise content of the class $\mathfrak{P}$, but on the other hand there is no need for such agreement, see Huskey [3]. For the purposes of this paper, a mapping $p\colon X \to Y$ is in the class $\mathfrak{P}$ if and only if there is a triangulation T of X such that $\Delta \varepsilon T$ implies 1° p is one-to-one on Δ, 2° $p(\Delta)$ is a rectilinear triangle in Y.

5.4. It is to be noted that p may be in $\mathfrak{P}$ by virtue of more than one triangulation of the above character. If $p \varepsilon P$, let $E(p) = \sum_{\Delta \varepsilon T} | p(\Delta) |$, where T is a triangulation of the type in **5.3**, and $| p(\Delta) |$ is the area of the triangle $p(\Delta)$. It is not difficult to see that $E(p)$ depends only on the mapping $p\colon X \to Y$ and is independent of the triangulation T. This number, $E(p)$, is called the *elementary* area of p.

5.5. The crucial point is the obvious one that $\mathfrak{P}$ is dense in $\mathscr{C}$, consequently if $f \varepsilon \mathscr{C}$, there is a sequence $\{p_n\}$ such that $p_n \to f$. With this particular sequence $\{p_n\}$ there is associated an element of P; namely $\liminf E(p_n)$. The infimum of these elements, $\liminf E(p_n)$, with respect to the totality of sequences $\{p_n\}$ such that $p_n \to f$ is again an element of P. It is this element which is defined to be $A(f)$; in short

$$A(f) = \inf_{p_n \to f} [\liminf_{n \to \infty} E(p_n)].$$

5.6. The non-trivial statements that $p \varepsilon \mathfrak{P}$ implies $A(p) = E(p)$, justifies speaking of $A(f)$ as an *area*, in fact, $A(f)$ will be called the Lebesgue area of f. An immediate consequence of this definition is that A is a lower semi-continous transformation.

5.7. It is now an easy matter to check that $(A, \tilde{L})$ is a coherent pair (**1.10**) in other words, that $f \tilde{L} g$ implies $A(f) = A(g)$. With but slight

additional effort it follows that $(A, \tilde{F})$ has the same property. On the other hand, the proof that $(A, \tilde{K})$ is a coherent pair is decidedly involved, see Radó [7], Helsel [2] and Youngs [13].

The fact that $\mathcal{L}$ cannot be properly topologized (**2.** 2) is unfortunate though it is possible to define a transformation $A_L\colon \mathcal{L} \to P$ using the device of **1.** 10. The fact that $\mathcal{F}$ and $\mathcal{K}$ can be properly topologized (**3.** 4 and **4.** 18) gives added content to the transformations $A_F\colon \mathcal{F} \to P$ and $A_K\colon \mathcal{K} \to P$ defined as in **1.** 10, for these transformations are known to be lower semi-continuous by the general theory of that paragraph. The number $A_F(\boldsymbol{F})$ is known as the Lebesgue area of the Fréchet surface $\boldsymbol{F}$, while $A_K(\boldsymbol{K})$ may be called the Lebesgue area of the Kerékjártó surface $\boldsymbol{K}$.

5. 8. Other things being equal, the general remarks of the first section, in particular **1.** 11, indicate that Kerékjártó equivalence is to be preferred to Fréchet equivalence simply because $F \cap K \neq 0$ implies $F \subset K$ and hence the class K may contain a representative with a particularly desirable property which is enjoyed by no representative of the class F. On the other hand, in the study of surfaces, there is good reason to continue to deal with Fréchet surfaces rather than change to those of Kerékjártó. To begin with, the proper topology on $\mathcal{F}$ makes it metric and while the same may be true of $\mathcal{K}$, no definite information is available on this point. The fact that $(A, \tilde{K})$ is a coherent pair should, for esthetic reasons, come early in the theory—as things stand, this is not the case (see **5.** 7). There is also another objection. Suppose $\boldsymbol{K}$ is a Kerékjártó surface, the space X being a closed 2-cell. There are strong analytic reasons for requiring that if f and g are representations of $\boldsymbol{K}$ (that is, f and g are mappings in K) then $f(\dot{X}) = g(\dot{X})$ where $\dot{X}$ is the boundary of X. Under the definition of Kerékjártó equivalence however, it is entirely possible for $f \tilde{K} g$ but $f(\dot{X}) \cap g(\dot{X}) = 0$ as Radó [7] has shown. Thus for quite compelling analytic reasons it would be necessary, in case X is a 2-cell, to modify Kerékjártó equivalence so as to generate the above property on the boundary of X. This can be done in a perfectly obvious manner, but it does not appear to be a particularly fruitful step. These statements may, by omission, give the impression that for Kerékjártó surfaces where X is a 2-sphere there is no corresponding analytic objection. On the contrary, if $\boldsymbol{K}$ is such a Kerékjártó surface then there is at least one important function defined over the representations of $\boldsymbol{K}$ which should be, but is *not* independent of the representation. The function referred to is $V(f)$, the volume enclosed by a mapping from a 2-sphere into Euclidean 3-space (see Radó [8]). Suppose that Z is a geometric 2-sphere in Y; let S be the bounded component of $Y - Z$ and $|S|$ its 3-dimensional Lebesgue measure. Suppose

that $\mathfrak{M}$ is a space consisting of two tangent 2-spheres. Let X be a 2-sphere, and select orientations on X and Z. It is easy to see that there are two monotone mappings, $m_1: X \rightrightarrows \mathfrak{M}$ and $m_2: X \rightrightarrows \mathfrak{M}$, and a light mapping $l: \mathfrak{M} \rightrightarrows Z$ such that Dgr $(lm_1) = 0$ and Dgr $(lm_2) = 2$. If $f_i = lm_i$, $i = 1, 2$, then $V(f_1) = 0$ while $V(f_2) = 2 \mid S \mid$. On the other hand, f_1 and f_2 are Kerékjártó equivalent. Such a situation does not arise for Fréchet equivalent mappings.

5. 9. All in all, therefore, the analytic aspects of surface theory indicate that Fréchet equivalence occupies a happy middle ground between Lebesgue and Kerékjártó equivalences. It is a fortunate property too, that there is a certain amount of inevitability about the concept—an equivalence relation at least as extensive as Lebesgue equivalence appears mandatory, and now if proper topologization of the resulting class of varieties is required, Fréchet equivalence (at least) is inevitable (**2.** 3).

BIBLIOGRAPHY.

1. M. Fréchet, "Sur la distance de deux surfaces," *Annales de la Société Polonaise de Mathématique*, vol. 3 (1924), pp. 4-19.
2. R. G. Helsel, "A theorem on surface area," *Transactions of the American Mathematical Society*, vol. 51 (1947), pp. 443-453.
3. H. D. Huskey, "Fréchet polyhedra," *Duke Mathematical Journal*, vol. 11 (1944), pp. 417-425.
4. B. V. Kerékjártó, "Involutions et surfaces continues," *Acta Litterarum ac Scientiarum* (Szeged), vol. 3 (1927), pp. 49-67.
5. H. Lebesgue, "Intégrale, longueur, aire," *Annali di Mathematica Pura ed Applicata*, vol. 7 (1902), pp. 231-359.
6. T. Radó, "On the semi-continuity of double integrals in parametric form," *Transactions of the American Mathematical Society*, vol. 51 (1942), pp. 336-361.
7. T. Radó, "On continuous mappings of Peano spaces," *Transactions of the American Mathematical Society*, vol. 58 (1945), pp. 420-454.
8. T. Radó, "The isoperimetric inequality and the Lebesgue definition of surface area," *Transactions of the American Mathematical Society*, vol. 61 (1947), pp. 530-555.
9. T. Radó and J. W. T. Youngs, "On upper semicontinuous collections," *Acta Litterarum ac Scientiarum* (Szeged), vol. 9 (1940), pp. 239-243.
10. G. T. Whyburn, "Regular convergence and monotone transformations," *American Journal of Mathematics*, vol. 57 (1935), pp. 902-906.
11. G. T. Whyburn, *Analytic Topology*, New York (1942).
12. J. W. T. Youngs, "Curves and surfaces," *American Mathematical Monthly*, vol. 51 (1941), pp. 1-11.
13. J. W. T. Youngs, "The topological theory of Fréchet surfaces," *Annals of Mathematics*, vol. 45 (1944), pp. 753-785.

SOME CHARACTERIZATIONS OF ARCS AND SIMPLE CLOSED CURVES.*

By R. H. BING.

1. Introduction. This paper uses the notion of *cuttings* to give some characterizations of arcs and simple closed curves. We make a distinction between *cutting* and *separating* as given in the following definitions:

DEFINITIONS. The set R *separates* A from B in the connected set M if $M - M \cdot R$ is the sum of two mutually separated sets containing A and B respectively. If M contains a continuum intersecting both A and B and R is a set containing neither A nor B but intersecting each continuum in M intersecting both A and B, we say that R cuts A from B in M. If a set separates (or cuts) two points from each other in M, we say that it separates (or cuts) M.

A *continuum* is a closed and connected set. It is *irreducible from H to K* if it intersects H and K but contains no proper subcontinuum doing so. It is *irreducible about R* if it contains R but none of its proper subcontinua does.

The theorems of this paper hold in a metric space. In fact, they hold in a Moore space, that is in a space satisfying Axiom 0 and the first three parts of Axiom 1 as given in R. L. Moore's *Foundations of Point Set Theory.*

J. R. Kline has shown [3] that for the plane, a nondegenerate continuum is a simple closed curve if it remains connected on the omission of any connected subset. This characterization is similar to the definition by R. L. Moore which defines [5] a simple closed curve to be a nondegenerate compact continuum which is separated by each pair of its points. However, the separating of a continuum by a pair of its points is not the same as the cutting of a continuum by a pair of its points. Every pair of points of a compact indecomposable continuum cuts it but no pair of points separates it. Therefore, the above mentioned characterizations by Kline and Moore differ from Theorem 11 of this paper which states that a nondegenerate compact continuum is a simple closed curve if it is cut by no one but by each pair of its points.

C. Kuratowski has shown [4] that if no subcontinuum of a nondegenerate bounded continuum C cuts it, then C is a simple closed curve. One of the results (Theorem 10) of this paper proves that instead of assuming that no

* Received September 25, 1947; Presented to the American Mathematical Society, November 28, 1947.

subcontinuum of C cuts C, we need only assume that no point cuts C and no subcontinuum of C separates it.

If a point separates a continuum, it separates two open subsets of the continuum from each other. Hence, to assume that a point cuts a continuum between two open subsets is a weaker hypothesis than to suppose that it separates the continuum. Some known results may be strengthened by replacing the notion of *separation* by the idea of *cutting between two open subsets.* For example, Corollary 2 is stronger than the known result that every compact continuum contains two points neither of which separates it and Theorem 9 is stronger than the known result that a compact continuum is a simple closed curve if every pair of its points separates it.

2. Cuttings. The five theorems of this section do not depend on each other. Most of them are used in later sections.

THEOREM 1. *Suppose that a, b, and c are three points of the continuum M which is irreducible from a to b. A necessary and sufficient condition that c separate a from b in M is that for each point p of $M - c$ there be a continuum in $M - p$ containing c and intersecting $a + b$.*

Proof. We shall prove only the sufficiency case of this theorem.

Suppose that for each point p of $M - c$ there is a continuum in $M - p$ containing c and intersecting $a + b$. Let X be the set of all points p such that some continuum in $M - p$ contains $c + b$ and let Y be the set of all points p such that some continuum in $M - p$ contains $c + a$. Each point of $M - c$ belongs to $X + Y$. We note that X contains a and Y contains b. Since M is irreducible from a to b, no point belongs to both X and Y.

For each point p of X there is a continuum in $M - p$ containing $c + b$ and there is a domain containing p but no point of the continuum. Each point of M in this domain is a point of X. Hence, X is open in M. Also, Y is open in M. Then X and Y are mutually separated sets. Since c separates X from Y, it separates a from b.

Example. Theorem 1 would not be true if instead of assuming that for each point p of $M - c$ there is a continuum in $M - p$ containing c and intersecting $a + b$, we only assumed that no point of $M - (a + b + c)$ cuts c from both a and b in M. To demonstrate this, let M be the closure of the graph in polar coordinates of $\rho = 1 + 1/\theta$ $(\theta \geq \pi)$; let a, b, and c be the points whose polar coordinates are $(1 + (1/\pi), \pi)$, $(1, \pi)$, and $(1, 0)$.

We shall make use of the following definition:

DEFINITION. A set M is aposyndetic [1] at the point p if p belongs to M and for each point x of M distinct from p, there exists an open subset of M which contains p and belongs to a connected and relatively closed subset of M lying in $M - x$.

THEOREM 2. *The compact continuum M which is not separated by any of its subcontinua is locally connected at the point p if it is aposyndetic at p.*

Proof. Let D be an open subset of M containing p. For each point q of $M - p$ there is a continuum $C(q)$ lying in $M - q$ and containing an open subset of M which contains p. Denote $M - C(q)$ by $D(q)$. Since $M - D$ is compact, there are a finite number of points $q_1, q_2, \cdots, q_n$ such that $D(q_1) + D(q_2) + \cdots + D(q_n)$ covers $M - D$.

Since no continuum separates M, $D(q_i)$ is connected and its closure $\bar{D}(q_i)$ is a continuum in $M - p$. We shall show that p does not cut $\bar{D}(q_i)$ from $\bar{D}(q_j)$ in M. Assume that this is not the case. Then $M - [\bar{D}(q_i) + \bar{D}(q_j)]$ is not a connected set, for if it were, the closure of this connected set would be a continuum in M separating q_i from q_j. Let $M - [\bar{D}(q_i) + \bar{D}(q_j)]$ be the sum of the mutually separated sets X and Y where Y contains p. Since neither $\bar{D}(q_i)$ nor $\bar{D}(q_j)$ separates M, $\bar{D}(q_i) + X + \bar{D}(q_j)$ is a continuum. Hence, p does not cut $\bar{D}(q_i)$ from $\bar{D}(q_j)$.

Since p does not cut any two of the continua $\bar{D}(q_1), \bar{D}(q_2), \cdots, \bar{D}(q_n)$ from each other in M, there is a continuum E in $M - p$ containing the sum of $\bar{D}(q_1), \bar{D}(q_2), \cdots$, and $\bar{D}(Q_n)$. The complement of E is a connected open subset of D containing p. This demonstrates that M is locally connected at p.

THEOREM 3. *If D and E are mutually exclusive open subsets of the compact continuum M and R is a subset of $M - (D + E)$ which cuts D from E in M, then the sum of the components of $M - (D + E)$ which intersect $\bar{R}$ separates D from E in M.*

Proof. Let $E_1, E_2, \cdots$ and $D_1, D_2, \cdots$ be sequences of open subsets of M such that $\Sigma E_i = E$, $\Sigma D_i = D$, D_{i+1} contains $\bar{D}_i$, and E_{i+1} contains $\bar{E}_i$. Then M is the sum of three mutually exclusive sets X_i, Y_i, and Z_i where Y_i is the sum of all components of $M - (D_i + E_i)$ intersecting $\bar{R}$, X_i is the sum of D_i and all components of $M - (D_i + E_i)$ that intersect D but not $\bar{R}$, and Z_i is the sum of E_i and all components of $M - (D_i + E_i)$ that intersect E but not $\bar{R}$. Now ΠY_i is the sum of the components of $M - (D + E)$ that intersect $\bar{R}$ and $M - \Pi Y_i$ is the sum of the mutually separated sets ΣX_i and ΣZ_i which contain D and E respectively.

COROLLARY 1. *If a point cuts between the open subsets D and E of the compact continuum M, then a continuum in $M - (D + E)$ separates D from E in M.*

THEOREM 4. *If D and E are mutually exclusive open subsets of the compact continuum M and R is a subset of $M - (D + E)$ which cuts $\bar{D}$ from $\bar{E}$ in M, then the sum of the components of $M - (D + E)$ which intersect R separates D from E in M.*

Proof. Let N be the sum of the components of $M - (D + E)$ which intersect R and X be the set of all points p of $M - N$ such that some continuum in $M - N$ intersects both $\bar{D}$ and p. We shall show that $M - N$ is the sum of two mutually separated sets one of which is X.

Now $\bar{X}$ is a subset of $X + N$; for let $p_1, p_2, \cdots$ be a sequence of points of X converging to a point p and denote the component of $M - (D + E)$ containing p_i by C_i. The limiting set C of $C_1, C_2, \cdots$ is a continuum intersecting both p and $\bar{D}$. Now C is a subset of N or X according as it does or does not intersect N. Likewise, the closure of $M - (N + X)$ is a subset of $M - X$. Therefore N separates X from $M - (N + X)$ in M and hence separates D from E.

THEOREM 5. *For each proper subset R of the compact continuum M there is a point p of $M - R$ such that the sum of all continua that lie in $M - p$ and intersect R is dense in M.*

Proof. Let N be a subcontinuum of M irreducible about R. First consider the case where $M = N$. Let p be any point of $N - R$. Assume that N contains an open subset E of itself such that each subcontinuum of N intersecting both $\bar{E}$ and R contains p. Then each component of $N - E$ which contains a point of R also contains p. Then there is a continuum in $N - E$ containing R contrary to the assumption that N is irreducible about R.

Next, consider the case where $M \neq N$. Since M is completely separable, there is a sequence of open subsets $D_1, D_2, \cdots$ of $M - N$ such that if D is an open subset of $M - N$ containing a point q, then there is an integer i such that D_i contains q and is a subset of D.

If an open subset of a proper subcontinuum of M contains $N + D_1$, let N_2 be such a subcontinuum; otherwise, let N_2 be N. If some proper subcontinuum of M has an open subset which contains $N_2 + D_2$, let N_3 be such a subcontinuum; otherwise, let N_3 be N_2. Similarly, we obtain $N_4, N_5, \cdots$. If ΣN_i is dense in M, a point of $M - \Sigma N_i$ is the required point p.

Suppose that ΣN_i is not dense in M. For convenience, suppose that $\Sigma N_i = N$. Then N does not belong to any open subset of any proper subcontinuum of M.

Let E_i be the component of $M - D_i$ containing N and F_i be the closure of $M - E_i$. Since $E_i \cdot F_i$ does not contain any open subset of M, the Theorem of Baire assures us that some point p of $M - N$ does not belong to $\Sigma E_i \cdot F_i$. If $F_i + N$ were the sum of two mutually separated sets X and Y where X contains N, N would belong to an open subset of the continuum $X + E_i$. Hence, $F_i + N$ is a continuum. Let G_i be the one of the continua E_i or $F_i + N$ that does not contain p. Although ΣG_i is dense in M, it does not contain p.

Corollary 2. *Each nondegenerate compact continuum contains two points neither of which cuts between two open subsets of the continuum.*

3. Characterizations of arcs. Using the facts that a compact continuum M is an arc between its points a and b if every point of $M - (a + b)$ separates a from b in M, we have the following theorem as an application of Theorem 1.

Theorem 6. *A necessary and sufficient condition that the compact continuum M which is irreducible from the point a to the point b be an arc from a to b is that for each pair of points p, q of M there is a continuum in $M - p$ containing q and intersecting $a + b$.*

It is to be noted that p is allowed to be either a or b. As is shown by the example following Theorem 1, the result would not be true if we restricted the range of p to $M - (a + b)$.

Theorem 7. *The compact continuum M is an arc between its points a and b provided each pair of points of $M - (a + b)$ cuts a from b in M and cuts an open subset of M from $a + b$ in M.*

Proof. Since M is irreducible from a to b, a does not cut M or else some two points which it cuts from b would not cut an open subset of M from b. Also, b does not cut M. As neither a nor b cuts M, we have by Theorem 6 that M is an arc if no point of $M - (a + b)$ cuts any point from $a + b$ in M. We shall show that the assumption that $M - (a + b)$ contains points p and q such that p cuts q from $a + b$ leads to the contradiction that $p + q$ does not cut any open subset of M from $a + b$.

Suppose that $p+q$ cuts the open subset D of M from $a+b$. Let ab be an arc from a to b such that $ab \cdot M = a+b$. The arc ab is not necessarily in the space containing M. By Theorem 3 there are continua H and K in $M-D$ such that H contains p, K contains q, X contains ab, and $M+ab$ is the sum of two closed proper subsets X and Y such that $X \cdot Y = H+K$.

Let F and G be the components of $M \cdot X$ containing a and b respectively. Each intersects $H+K$ but not both intersect the same one of H, K or else some pair of points of Y does not cut a from b in M. But p does not cut q from a or from b according as K intersects F or G.

THEOREM 8. *The compact continuum M is an arc between its points a and b provided each pair of points of $M-(a+b)$ cuts an open subset of M from an open subset of M containing $a+b$.*

Proof. Assume first that some subcontinuum N of M is irreducible from a to b and is not an arc. Then by the preceding theorem, $N-(a+b)$ contains two points p and q whose sum does not cut any open subset of N from $a+b$ in N. However, $p+q$ cuts between two open subsets D and E of M in M where E does not intersect N and D contains $a+b$.

By Theorem 3, there are two continua H and K in $M-(D+E)$ whose sum separates D from E in M and such that H contains p and K contains q. Suppose that $M-(H+K)$ is the sum of two mutually separated sets $S(D)$ and $S(E)$ which contain D and E respectively. By Theorem 5 there is a point r of $S(E)-S(E)\cdot N$ that does not cut any open subset of $S(E)$ from $H+K+N$ in M.

If $S(D)$ is not a subset of N, by Theorem 5 there is a point t of $S(D)-S(D)\cdot N$ that does not cut any open subset of $S(D)$ from $H+K+N$ in M. Then $r+t$ does not cut any open subset of M from N and hence does not cut any open subset of M from $a+b$.

We shall show that in case N contains $S(D)$, there is a point t of $S(D)$ that does not cut either H or K from any domain containing $a+b$ in $N+H+K$. Then $r+t$ will not cut any open subset of M from an open subset of M containing $a+b$.

If N were indecomposable, we could use any point t in $S(D)$ belonging to a composant of N not intersecting $a+b+p+q$. Hence, suppose that it is decomposable. Then either a or b does not cut p from $a+b$ in N. Also, either a or b does not cut q from $a+b$. If a (or b) cuts neither p nor q from b (or a), then there is a domain containing a (or b) such that any

point of $S(D) - a$ (or $S(D) - b$) in this domain will serve as t. Hence, suppose that a cuts p from b and b cuts q from a in N.

Let $D_1, D_2, \cdots$ be a sequence of open subsets of N closing down on a where D_1 does not intersect $p+q$ and let C_{p_i} and C_{q_i} be the components of $N - D_i$ containing p and q respectively. C_{p_i} does not intersect C_{q_i} for any integer i because there is an integer j such that C_{q_j} contains b. Let F be an open subset of $S(D)$ containing a but no point of C_{q_j}. Since the limiting set of $C_{q_1}, C_{q_2}, \cdots$ is n, there is an integer k such that C_{q_k} intersects F. Then no point t of $F \cdot C_{q_k}$ cuts q from b and p from any domain containing a in N.

We have shown that the assumption that there is a subcontinuum of M irreducible from a to b which is not an arc leads to the contradiction that some pair of points r, t of $M-(a+b)$ does not cut an open subset of M from an open subset of M containing $a+b$. Hence, every subcontinuum of M irreducible from a to b is an arc. Let ab be one such arc. We shall show that if M is not ab, then there are two points of $M-(a+b)$ whose sum does not cut any open subset of M from any open subset of M containing $a+b$.

Let X be the collection of all continua x such that x is the closure of a maximal continuumwise connected subset of $M - ab$. The sum of a countable subcollection of X is dense in $M - ab$. If this countable subcollection failed to cover two points of $M - ab$, this pair of points would not cut an open subset of M from ab in M.

Since a countable subcollection of X covers $M - ab$, some element x_1 of X contains an open subset of $M - ab$. First, consider the case where x_1 covers $M - ab$. If x_1 contains ab, no pair of points of $ab - (a+b)$ would cut an open subset of M from an open subset of M containing $a+b$ for, if so, each open subset of M would intersect the maximal continuumwise connected subset of $M - ab$ whose closure is x_1. If r is a point of $ab - (a+b)$ not belonging to x_1 and t is a point of $M - ab$ not cutting any open subset of M from ab (Theorem 5) then $r+t$ would not cut any open subset of M from $a+b$ in M.

If x_1 does not cover $M - ab$, there is another element x_2 of X which contains an open subset of $M - ab$. Let C_i $(i=1,2)$ be the closure of the sum of ab and all elements of X other than x_i. By Theorem 5 there is a point p_i of $M - C_i$ that does not cut any open subset of M from C_i in M. Any subcontinuum of M irreducible from p_i to ab is a subset of x_i. Then $p_i + p_2$ do not cut any open subset of M from ab in M. Hence, M is ab.

4. Characterizations of a simple closed curve. We shall now use cuttings of continua to give some characterizations of a simple closed curve.

THEOREM 9. *The compact continuum M is a simple closed curve if each pair of its points cuts between two open subsets of M.*

Proof. No subcontinuum N of M separates M because if $M - N$ were the sum of two mutually separated sets X and Y, then by Theorem 5 there would be points p and q in X and Y respectively such that continuumwise connected subsets of $N + X - p$ and $N + Y - q$ would contain N and be dense in $N + X$ and $N + Y$ respectively. Then $p + q$ would not separate any two open subsets of M from each other.

If the points p and q cut between two open subsets, then by Theorem 3 there are subcontinua H and K of M such that $M - (H + K)$ is the sum of two mutually separated sets X and Y. Since no subcontinuum of M separates it, $\bar{X}$ is ireducible from H to K or else the sum of H, K, and a proper subcontinuum of $\bar{X}$ is a continuum separating M. By Theorem 7 we find that every point of X separates H from K in $H + X + K$. Then some two points separate M. Reapplying Theorem 7, we find that M is the sum of two arcs between these points and is in fact a simple closed curve.

THEOREM 10. *A nondegenerate compact continuum is a simple closed curve if it is neither cut by any point nor separated by any one of its subcontinua.*[1]

Proof. Let M be such a continuum. It has been shown [2] by F. B. Jones that a compact continuum which is not cut by any of its points is aposyndetic on a dense subset. Using this result and Theorem 2, we find that M is locally connected at two of its points a, b.

Let $D_1, D_2, \cdots$ and $E_1, E_2, \cdots$ be sequences of connected domains closing down on a and b respectively such that D_1 does not intersect E_1. Now $M - (D_i + E_i)$ is the sum of two mutually separated sets X_i and Y_i, for if it were connected. it would be a continuum separating a from b in M. No proper subcontinuum of $\bar{D}_i + X_i + \bar{E}_i$ contains both D_i and E_i or else a subcontinuum of M separates a point of X_i from Y_i. Also $\bar{Y}_i$ is irreducible from $\bar{D}_i$ to $\bar{E}_i$. Suppose that X_{i+1} intersects X_i. Then X_i is a subset of X_{i+1}. Now M is the sum of two continua $a + \Sigma X_i + b$ and $a + \Sigma Y_i + b$ and it follows from Theorem 6 that each of these is an arc from a to b.

---

[1] By modifying our arguments we could have strengthened Theorems 2, 6, 10, 11, and 12 by weakening the hypotheses of these theorems by supposing the sets to be locally peripherally compact instead of compact.

THEOREM 11. *A nondegenerate compact continuum is a simple closed curve if it is cut by no one but by each pair of its points.*

Proof. Suppose that M is such a continuum. It will follow from Theorem 10 that M is a simple closed curve if we can show that none of its subcontinua separates it.

Assume that M contains a continuum H such that $M - H$ is the sum of two mutually separated sets X and Y containing points p and q respectively. We shall show that this assumption leads to the contradiction that $p + q$ does not cut any point of M from H in M.

Assume that $p + q$ cuts the points a of X from H in M. Since no point cuts M, there is a continuum C in $M - p$ containing $a + H$. Now $C - C \cdot Y$ is a closed set in $M - (p + q)$ containing $a + H$. Also, it is a continuum, for if it were the sum of the mutually separated sets R and S where S contains H, then C would be the sum of the mutually separated sets R and $S + C \cdot Y$. Also, $p + q$ does not cut any point of Y from H and therefore does not cut M.

THEOREM 12. *A nondegenerate compact connected space which is cut by no one of its points is a simple closed curve if each of its proper subcontinua is cut by each interior point of this subcontinuum.*

Proof. Assume that the complement of the continuum H is the sum of two mutually separated sets X and Y and let p be a point of X. Then p does not cut the continuum $H + X$ because it does not cut the space. Hence, no continuum separates the space and Theorem 12 follows from Theorem 10.

THEOREM 13. *The nondegenerate compact continuum M which is cut by no one of its points is a simple closed curve provided that for each three points a, b, p of M there is a point q such that $p + q$ cuts a from b in M.*

Proof. Again we show that no subcontinuum H of M separates it. Assume that $M - H$ is the sum of two mutually separated sets X and Y. Let a and b be two points of a continuum K in X and let p be a point of Y. If there were a point q such that $p + q$ cuts a from b, then q belongs to K. But there is a continuum in $M - q$ containing $a + H$ and another continuum in $M - q$ containing $b + H$. The sum of these two continua is a continuum in $M - (p + q)$ containing $a + b$. Hence, no subcontinuum separates M.

Example. We give an example to show that a compact continuum which is cut by no point need not be a simple closed curve even though for each pair of its points there is another pair of points cutting between these two in the continuum. Let the continuum be the sum of the graph of $y = \sin(1/x)$ $(0 < x \leqq 4)$ and the square with opposite vertices at $(0, 2)$ and $(4, -2)$.

THE UNIVERSITY OF WISCONSIN.

REFERENCES.

1. F. B. Jones, "Aposyndetic continua and certain boundary problems," *American Journal of Mathematics*, vol. 63 (1941), pp. 545-553.

2. ———, "Concerning non-aposyndetic continua," *American Journal of Mathematics*, vol. 70 (1948), pp. 403-413.

3. J. R. Kline, "Closed connected sets which remain connected upon the removal of certain connected subsets," *Fundamenta Mathematicae*, vol. 5 (1924), pp. 3-10.

4. C. Kuratowski, "Contribution à l'étude de continus de Jordan," *Fundamenta Mathematicae*, vol. 5 (1924), pp. 112-127.

5. R. L. Moore, "Concerning simple continuous curves," *Transactions of the American Mathematical Society*, vol. 21 (1920), pp. 333-347.

ON THE DIMENSION OF PARTIALLY ORDERED SETS.* †

By Horace Komm.

1. **Introduction.** A *partial order* P is a set of elements S together with a binary relation $R(x,y)$ which may hold between any two elements x and y of S, and which satisfies the two conditions of asymmetry and transitivity; i. e. (1) if $R(x,y)$ then not $R(y,x)$, and (2) if $R(x,y)$ and $R(y,z)$ then $R(x,z)$.

In what follows we read $R(x,y)$ as "x precedes y" and write "$x < y$," or alternately as "y follows x" and "$y > x$." Two distinct elements x and y of S are comparable in P if either $x < y$ or $y < x$. Otherwise x and y are non-comparable, and we will write "$x \phi y$." If x and y are the same element of S we will write $x = y$. From (1) it follows that $R(x,x)$ never holds.

The most familiar example of a partial order is that furnished by the set of all subsets of a given set with set inclusion as the ordering relation. It may be mentioned here that, conversely, any partial order is similar to a family of sets ordered by set inclusion. For, if P is defined on S, let $S(x)$ consist of x and all $y \varepsilon S$ such that $y < x$ in P. It is easily seen that the partial order P' defined on $\{S(x)\}$ for all x in S by means of set inclusion is similar to P, where the correspondence is between x and $S(x)$. By *similarity* between P defined on S and P' defined on S' is meant the existence of a $1-1$ correspondence between S and S' which is order preserving; i. e. $x' < y'$ in P' if and only if $x < y$ in P, where x' and y' are the elements of S' corresponding to x and y of S. If P is similar to P' we shall write $P \sim P'$.

A partial order P defined on a set S is a *linear order* if every distinct pair of elements of S is comparable; i. e. if $R(x,y)$, in addition to (1) and (2), satisfies (3) if $x \neq y$, then either $R(x,y)$ or $R(y,x)$. If P is a partial order on a set S, and L is a linear order on S, then L will be called a *linear extension* of P provided $x < y$ in L whenever $x < y$ in P. For example, let S be the set of points in the plane given by $0 \leq x \leq 1$, $0 \leq y \leq 1$, P the partial order defined on S by the condition that $(x_1, y_1) < (x_2, y_2)$ if and only if $x_1 < x_2$, and L the linear order defined on S by the condition that

† Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the University of Michigan. I would like to express here my very great indebtedness to Professor Dushnik and the late Professor E. W. Miller.

* Received July 5, 1947.

$(x_1, y_1) < (x_2, y_2)$ in L if $x_1 < x_2$ or if $x_1 = x_2$ and $y_1 < y_2$. Clearly, L is a linear extension of P.

Now, let $\mathcal{K}$ be a collection of linear orders on a set S, and P the partial order on S obtained by putting $x < y$ in P if and only if $x < y$ in every linear order of $\mathcal{K}$. The partial order P thus defined is said to be *realized* by $\mathcal{K}$.

An immediate consequence of this definition is that if P is a partial order realized by a collection $\mathcal{K}$ of linear orders, then, for every non-comparable pair of elements x and y of P, there exists a linear order L_1 of $\mathcal{K}$ in which $x < y$, and a second linear order L_2 of $\mathcal{K}$ in which $y < x$.

More generally, let S be any set, and $\{L_\alpha\}$ $(\alpha < \beta)$ a series of linear orders, each L_α being defined on a set S_α. Let $\{f_\alpha\}$ $(\alpha < \beta)$ be a series of single-valued functions, each with a single-valued inverse, such that $f_\alpha(S) \subset S_\alpha$. We define the partial order P on S by specifying for any two elements x and y of S that $x < y$ in P if and only if $f_\alpha(x) < f_\alpha(y)$ for all $\alpha < \beta$. If P is defined in this way it is realized by $\{f_\alpha\}$.

Finally, we define the *dimension* of a partial order P as the smallest cardinal number $\mathfrak{a}$ such that P is realized by $\mathfrak{a}$ linear extensions [1].[1] The dimension of P will be denoted by $\dim P$.

The following results have already been obtained.

1.1. THEOREM [2]. *Every partial order P possesses a linear extension L. Moreover, if x and y are any two non-comparable elements of P, there exists an extension L_1 in which $x < y$ and an extension L_2 in which $y < x$.*

1.2. THEOREM [1]. *If P is any partial order on a set S, then there exists a collection $\mathcal{K}$ of linear orders on S which realize P.*

1.3. THEOREM [1]. *Let P be any partial order on a set S. If S is finite, then the dimension of P is finite. If $\bar{\bar{S}} \doteq \mathfrak{m}$, where $\bar{\bar{S}}$ is the cardinality of S and $\mathfrak{m}$ is a transfinite cardinal, then the dimension of P is $\leqq \mathfrak{m}$.*

1.4. THEOREM [1]. *For every cardinal number $\mathfrak{m}$ (finite or transfinite), there exists a partial order, defined on a set of power $2\mathfrak{m}$, whose dimension is $\mathfrak{m}$.*

In **3**.1 we define in an obvious way two partial orders (denoted by P_n and P'_n) on the set of all finite or denumerable sequences (depending on n). With the help of these the following main results are obtained: (1) Every finite or denumerable partial order is similar to some subset of P_n (or P'_n). (2) By a modification of the definition of the dimension of a partial order

[1] Numbers in brackets refer to the references cited at the end of the paper.

(see **4**) it is shown that this result applies also to the wider class of partial orders whose λ-dimension is $\leq \aleph_0$. (3) $\dim P_n = \dim P'_n = n$, n finite or $\aleph_0$; λ-$\dim P'_n = \aleph_0$, λ-$\dim P_n = c$, for finite n, where c is the power of the continuum. (4) A characterization of those subsets of the euclidean n-space for which P_n and P'_n have finite λ-dimension is obtained. Finally, we have the results somewhat unrelated to the rest, that the dimension of the partial order $P(\mathfrak{a})$ on the set of all subsets of a set of power $\mathfrak{a}$ is $\mathfrak{a}$, and λ-$\dim P(\aleph_0) = \aleph_0$.

2. The dimension of $P(\mathfrak{a})$. A simple example of a partial order (which is also a lattice) is provided by the set of all subsets $\mathfrak{M}$ of any set M, with set inclusion as the ordering relation. Denote by $P(\mathfrak{a})$ the partial order defined on the set of all subsets of any set of power $\mathfrak{a}$ (finite or transfinite) with respect to the relation of set inclusion. We have the following theorem.

2.1. THEOREM. *The dimension of $P(\mathfrak{a})$ is $\mathfrak{a}$.*

Proof. Let $P(\mathfrak{a})$ be defined on $\mathfrak{M}$, the set of all subsets of a set M of power $\mathfrak{a}$. For each x in M, denote by a_x the subset of M whose only element is x. Let $\mathfrak{M}'_x$ be the set of subsets of M which are non-comparable with or less than (of the latter there will be only the null set) a_x in $P(\mathfrak{A})$; and let $\mathfrak{M}''_x$ be the set of all subsets of M which are greater than a_x in $P(\mathfrak{a})$. Denote by P'_x the partial order (with respect to the relation of set inclusion) on $\mathfrak{M}'_x$, and by P''_x the partial order on $\mathfrak{M}''_x$. Finally, let L'_x be any linear extension of P'_x and L''_x any linear extension of P''_x. For each x in M, define the linear extension L_x of $P(\mathfrak{a})$ by specifying that $L'_x < a_x < L''_x$ in L_x. Consider the set $\{L_x\} = \mathcal{K}$ of linear extensions of $P(\mathfrak{a})$, where x runs over all the elements of M. Clearly, $\bar{\mathcal{K}} = \mathfrak{a}$. Further, $P(\mathfrak{a})$ is realized by $\mathcal{K}$. For, suppose A and A' to be any two subsets of M which are non-comparable in $P(\mathfrak{a})$. Then there exists an element x of A which does not belong to A', and an element y of A' which does not belong to A. Now, $a_x \leq A$ and $a_x \phi A'$ in $P(\mathfrak{a})$, so that $A' < a_x \leqq A$ in L_x; while $a_y \leqq A'$ and $a_y \phi A$ in $P(\mathfrak{a})$, so that $A < a_y \leqq A'$ in L_y. Therefore, $P(\mathfrak{a})$ is realized by $\{L_x\}$, and $\dim P(\mathfrak{a}) \leqq \mathfrak{a}$. On the other hand, let c_x and c_y denote the complements of a_x and a_y in M. Suppose that $c_x < a_x$ and $c_y < a_y$ in the same linear extension of $P(\mathfrak{a})$. Then $c_x < a_x < c_y < a_y$ which contradicts the fact that $a_y < c_x$ in $P(\mathfrak{a})$. It follows that in order to realize $P(\mathfrak{a})$ it is necessary to have a linear extension for each x in M, and therefore, the dimension of $P(\mathfrak{a})$ cannot be less than $\mathfrak{A}$. This proves the theorem.

3. The dimension of P_n and P'_n. Let E_n be the set of all sequences of real numbers $\{x_k\}$, $k = 1, 2, \cdots, n$ where $n \leq \aleph_0$. When n is finite we shall consider the sequences $\{x_k\}$ as being the points of the euclidean n-space, also denoted by E_n. This section deals with two partial orders, $P_n(M_n)$ and $P'_n(M_n)$, defined on subsets M_n of E_n. The significance of these partial orders from the point of view of this paper lies in the fact that every finite or denumerable partial order of dimension n is similar to $P_n(M_n)$ for some $M_n \subset E_n$. In **4** this result is strengthened to include every partial order which can be realized by an at most denumerable set of linear extensions each of which is similar to a subset of the linear continuum.

3.1. *Definition.* Let M_n be any subset of E_n. If $a = \{x_k\}$, $b = \{y_k\}$ are any two points of M_n put $a < b$ in $P_n(M_n)$ if and only if $x_k < y_k$ for all k; and put $a < b$ in $P'_n(M_n)$ if and only if $x_k \leq y_k$ for all k, and there exists an i such that $x_i < y_i$.

We shall now prove the following theorem.

3.2. THEOREM. *If P is a denumerable partial order of dimension n, there exists a subset M_n of E_n such that $P \sim P_n(M_n)$.*

Proof. Let P be a partial order of dimension n defined on a denumerable set S. It follows that $n \leq \aleph_0$ [Theorem 1.3]. Let $\{L_i\}$ be a sequence of linear orders on S which realize P. Since every denumerable linear order is similar to a subset of the rational numbers ordered with respect to magnitude, there exists a sequence of single-valued functions $\{f_i\}$, each with single-valued inverse, such that $f_i(S) \subset R$, where R is the set of rational numbers, and $f_i(a) < f_i(b)$ if and only if $a < b$ in L_i. We also have $f_i(a) < f_i(b)$ for all i if and only if $a < b$ in P since $\{L_i\}$ realizes P.

Now, define $\psi(S) = M_n \subset E_n$ by $\psi(a) = \{f_k(a)\}$ for $a \in S$. Then, if $a < b$ in P, $f_i(a) < f_i(b)$ for all i and $\psi(a) < \psi(b)$ in $P_n(M_n)$. If $a \phi b$ in P, then $f_i(a) < f_i(b)$ for some i, $f_j(b) < f_j(a)$ for some $j \neq i$, and $\psi(a) \phi \psi(b)$ in $P_n(M_n)$. It follows that $P \sim P_n(M_n)$.

As immediate consequences of Theorem 3.2 we have

3.21. COROLLARY. *If P is a finite partial order, then there exists a finite positive integer n, and a subset M_n of E_n such that $P \sim P_n(M_n)$.*

3.22. COROLLARY. *If P is a finite or denumerable partial order, there exists a subset M of the rational points of $E_{\aleph_0}$ such that $P_{\aleph_0}(M) \sim P$.*

3.23. COROLLARY. 3.2, 3.21, 3.22 *remain true when P'_n is substituted for P_n.*

We can now prove the following.

3. 3. THEOREM. *The dim* $P_n(E_n) = n$, *where* $n \leq \aleph_0$.

Proof. To prove that the dimension of $P_n(E_n)$ is at least n it is sufficient to show that $P_n(E_n)$ contains a partial order whose dimension is n. By 3. 2 and 3. 21, it is sufficient to show that, given any cardinal $n \leq \aleph_0$, there exists a finite or denumerable partial order having dimension n. This follows from 1. 4, and, therefore, $\dim P_n(E_n) \geq n$.

To complete the proof we exhibit n linear extensions of $P_n(E_n)$ which realize it. If a and b are any two distinct points of E_n, put $a < b$ in $L_n{}^i$ if (1) $x_i{}^{(a)} < x_i{}^{(b)}$; or if (2) $x_i{}^{(a)} = x_i{}^{(b)}$ and $x_j{}^{(a)} > x_j{}^{(b)}$, where j is the first subscript for which $x_j{}^{(a)} \neq x_j{}^{(b)}$. The order $L_n{}^i$ is linear. Obviously, every two points of E_n are ordered in $L_n{}^i$. To show that transitivity holds let a, b, c be any elements of E_n such that $a < b$, $b < c$ in $L_n{}^i$. If $x_i{}^{(a)} < x_i{}^{(b)}$ and $x_i{}^{(b)} \leq x_i{}^{(c)}$, or $x_i{}^{(a)} = x_i{}^{(b)}$ and $x_i{}^{(b)} < x_i{}^{(c)}$, the result is immediate. If $x_i{}^{(a)} = x_i{}^{(b)}$ and $x_i{}^{(b)} = x_i{}^{(c)}$ there exist $j, k \neq i$ such that $x_j{}^{(a)} > x_j{}^{(b)}$, $x_k{}^{(b)} > x_k{}^{(c)}$ and $x_l{}^{(a)} = x_l{}^{(b)}$, $x_s{}^{(b)} = x_s{}^{(c)}$ for $l < j$, $s < k$. It follows that if $j \leq k$, $x_j{}^{(a)} > x_j{}^{(c)}$ and $x_l{}^{(a)} = x_l{}^{(c)}$ for $l < j$; or if $j > k$, $x_k{}^{(a)} > x_k{}^{(c)}$ and $x_l{}^{(a)} = x_l{}^{(c)}$ for $l < k$. In either case $a < c$. Further, $\{L_n{}^i\}$ realizes $P_n(E_n)$. For, if $a < b$ in $P_n(E_n)$, $x_i{}^{(a)} < x_i{}^{(b)}$ for all i, and $a < b$ in $L_n{}^i$ for all i. If $a \phi b$ in $P_n(E_n)$, either (1) $x_i{}^{(a)} = x_i{}^{(b)}$ for some i, or (2) $x_k{}^{(a)} < x_k{}^{(b)}$ for some k and $x_l{}^{(a)} > x_l{}^{(b)}$ for some $l \neq k$. If (1), since $a \neq b$, there exists a first j such that either $x_j{}^{(a)} < x_j{}^{(b)}$ or $x_j{}^{(a)} > x_j{}^{(b)}$. If $x_j{}^{(a)} < x_j{}^{(b)}$, $a < b$ in $L_n{}^j$ and $a > b$ in $L_n{}^i$. If $x_j{}^{(a)} > x_j{}^{(b)}$, $a > b$ in $L_n{}^j$ and $a < b$ in $L_n{}^i$. If (2), $a < b$ in $L_n{}^k$ and $a > b$ in $L_n{}^l$. Therefore, $\{L_n{}^i\}$ realizes $P_n(E_n)$, and $\dim P_n(E_n) = n$.

3. 4. THEOREM. *The dim* $P'_n(E_n) = n$, $n \leq \aleph_0$.

The proof of 3. 3 applies here also if, for the realization of $P'_n(E_n)$, is chosen the series of linear orders $\{\bar{L}_n{}^i\}$ on E_n defined by putting $a < b$ in $\bar{L}_n{}^i$ if (1) $x_i{}^{(a)} < x_i{}^{(b)}$, or if (2) $x_i{}^{(a)} = x_i{}^{(b)}$ and $x_j{}^{(a)} < x_j{}^{(b)}$, where j is the first subscript for which $x_j{}^{(a)} \neq x_j{}^{(b)}$.

4. The λ-dimension of a partial order. In this section, in order to strengthen the results of the last section, we introduce the notion of the α-dimension of a partial order, and, as a particular case, the λ-dimension of a partial order.

4. 1. *Definition.* The *α-dimension* of a partial order P is the smallest cardinal number $\mathfrak{a}$ such that P is realized by $\mathfrak{a}$ linear extensions each of

which is similar to a subset of a set of order type α. Such linear extensions will be called α-extensions. In particular, when $\alpha = \lambda$, the order type of the linear continuum, the α-dimension will be called the *λ-dimension.*

What follows is restricted mostly to a consideration of the λ-dimension. An example of a partial order which does not have λ-dimension is furnished by the linear order defined on the ordinals of the first and second classes with respect to magnitude. An even more obvious example is any partial order defined on a set S whose power is greater than c, the power of the continuum.

The following theorem establishes a simple criterion for determining whether a given partial order has λ-dimension.

4.2. THEOREM. *In order that a partial order P defined on a set S have λ-dimension, it is necessary and sufficient that P have a λ-extension.*

Proof. The necessity of this condition is obvious. To prove the condition sufficient assume that P has a λ-extension K. If P contains no non-comparable elements the theorem is trivial. Therefore, let a and b be any two elements of S such that $a \phi b$ in P, and suppose $a < b$ in K. To prove the theorem it is sufficient to show that there exists a second λ-extension K' of P in which $a > b$.

Since K is a λ-extension of P, there is a single-valued function f, with single-valued inverse, such that $f(K) \subset L$, where L is the linear continuum. Let $f(a) = r_1$, $f(b) = r_2$, where r_1 and r_2 are real numbers and $r_1 < r_2$, and let

$$A = \underset{x}{E}[x \varepsilon S,\ a < x \text{ in } P \text{ and } r_1 < f(x) < r_2].$$

$$B = \underset{x}{E}[x \varepsilon S,\ x < b \text{ in } P \text{ and } r_1 < f(x) < r_2].$$

$$C = \underset{x}{E}[x \varepsilon S,\ x \phi a,\ x \phi b \text{ in } P \text{ and } r_1 < f(x) < r_2].$$

$$D = \underset{x}{E}[x \varepsilon S,\ f(x) < r_1].$$

$$E = \underset{x}{E}[x \varepsilon S,\ f(x) > r_2].$$

Clearly, $f(A) + f(B) + f(C) = f(S) \cdot (r_1, r_2)$, and $f(A) \cdot f(B) = f(B) \cdot f(C) = f(A) \cdot f(C) = 0$. Let k be a constant greater than $r_2 - r_1$, and define the function $f'(S) \subset L$ as follows: $f'(a) = r_2$, $f'(b) = r_1$, $f'(A) = f(A) + k$, $f'(B) = f(B) - k$, $f'(C) = f(C)$, $f'(D) = f(D) - k$, $f'(E) = f(E) + k$. Now, define the linear order K' on S by putting $x < y$ in K' if and only if $f'(x) < f'(y)$, where x and y are any two elements of S. Since $r_1 < r_2$, $f'(a) > f'(b)$, so that $a > b$ in K'. To show that K' is an extension of P,

consider any two elements p and q of S such that p is comparable to q in P. There are the following non-trivial cases: (1) $p \varepsilon C$, $q \varepsilon B$. Then $q < p$ in P, for otherwise $p < q < b$ in P. Therefore, $f(q) < f(p)$ and $f'(q) < f'(p)$, so that $q < p$ in K'. (2) $p \varepsilon C$, $q \varepsilon A$. Then $p < q$ in P, for otherwise $a < q < p$ in P. Therefore, $f(p) < f(q)$ and $f'(p) < f'(q)$, so that $p < q$ in K'. (3) $p \varepsilon A$, $q \varepsilon B$. Then $q < p$, and we have the same result as in (1).

Therefore, K' is a λ-extension of P in which $a > b$, and P has λ-dimension.

4.21. Corollary. *Any partial order P which has λ-dimension n (finite or $\aleph_0$) is similar to $P_n(M_n)$ for some $M_n \subset E_n$.*

4.22. Corollary. *If P is any finite or denumerable partial order, P has λ-dimension; and the dimension of P is the same as the λ-dimension of P.*

Proof. This is an immediate consequence of the fact that any linear extension L of P is similar to a subset of the rational numbers ordered with respect to magnitude and is, therefore, a λ-extension of P.

It should also be remarked here that the λ-dimension of a partial order, if it exists, is greater than or equal to the dimension of the partial order (see 4.5 for an example of a partial order whose λ-dimension is greater than its dimension).

A generalization of Theorem 4.2 may be stated on the basis of the following definition.

4.3. *Definition.* A linear order L of type α will be called *homogeneous* if either (1) every closed interval of L contains at most a finite number of elements, or (2) every closed interval of L contains a proper subset of order type α.

4.4. Theorem. *In order that a partial order P have α-dimension, where α is any homogeneous order type, it is necessary and sufficient that P have an α-extension.*

The proof of 4.2 with a slight modification is applicable here also. The modification consists of translating the sets A, B, C, D, E purely in terms of order, but this is possible because of the homogeneity of the order type α.

The question arises as to whether, given a partial order P whose λ-dimension exists, the λ-dimension is always equal to the dimension of P. The answer to this question is in the negative, as the following example shows.

4.5. *Example.* Let A be any set of power $\aleph_1$. Define the partial order P on A by two single-valued functions with single-valued inverses, f and f',

in the following way: (1) $f(A) \subset L$, where L is the linear continuum, (2) $f'(A) \subset W$, where W is the linear order, with respect to magnitude, of the ordinals of the first and second classes. The mapping f is possible since $\aleph_1 \leq c$, where c is the power of the continuum.

The dimension of P is 2, and P has λ-dimension by Theorem 4. 2. Suppose that the λ-dimension of P is n, for some finite n. Then, by Corollary 4. 21, $P \sim P_n(M_n)$ for some $M_n \subset E_n$. Since $\bar{\bar{M}}_n = \aleph_1$, M_n contains a non-denumerable set of condensation points. Denote the set of condensation points of M_n by M'_n. Now, $P_n(M'_n)$ is not chaotic; i. e., $P_n(M'_n)$ contains at least two comparable points $a, b \in M'_n$. Suppose $a < b$. Then $x_k^{(b)} - x_k^{(a)} > 0$, $k \leq n$. Choose $\epsilon = \frac{1}{2} \min (x_k^{(b)} - x_k^{(a)})$, and denote by $S(a, \epsilon)$ the interior of an open n-sphere of radius ϵ with a as center. Since a is a condensation point of M_n, $S(a, \epsilon)$ contains a non-denumerable subset $\bar{M}_n$ of M_n. If c is any point of $\bar{M}_n$, we have $x_k^{(c)} < x_k^{(a)} + \epsilon < x_k^{(a)} + \frac{1}{2} \min (x_k^{(b)} - x_k^{(a)}) < x_k^{(a)} + x_k^{(b)} - x_k^{(a)}$. Therefore, $\bar{M}_n < b$ in $P_n(M_n)$. It follows that the element p of A corresponding to b of M_n is preceded in P by a non-denumerable subset M of A, and, therefore, $f'(M) < f'(p)$. This, however, is impossible due to the nature of W, so that the λ-dimension of P is not finite and differs from the dimension of P.

5. The λ-dimension of $P_n(E_n)$ and $P'_n(E_n)$.

5. 1. THEOREM. *The λ-dimension of $P'_n(E_n)$ exists for every finite n.*

Proof. By Theorem 4. 2, it is sufficient to show that $P'_n(E_n)$ has a λ-extension. A λ-extension of $P'_n(E_n)$ may be thought of as a real-valued function, $f(x_1, \cdots, x_n)$, of n real variables, which is monotonic increasing in each variable, and which is single-valued and has a single-valued inverse. We exhibit such a function.

Consider the set of values $[\psi]$ of the function $\psi(x)$ defined by

$$\psi(x) = \sum_{n=0}^{\infty} \frac{2^{\alpha_n}}{2^{\beta_n}} \quad ; \quad \alpha_n = 2^{Enx}; \quad \beta_n = 2^{n^2}$$

for $x > 0$, where Enx is the largest integer $\leq nx$. This set was defined by von Neumann [3, p. 135] and has the following properties which are useful in the present context: (1) Any finite subset of $[\psi]$ consists of numbers which are algebraically independent, i. e., $F(a_1, \cdots, a_k) = 0$ if and only if $F(a_1, \cdots, a_k)$ is identically 0, where $F(a_1, \cdots, a_k)$ is any polynomial in $a_1, \cdots, a_k \in [\psi]$ with rational coefficients [3, p. 136]. (2) $\psi(x)$ is monotonic increasing [4, p. 17]. (3) $[\psi]$ contains a perfect subset K [4, p. 18].

Let $K_1, \cdots, K_n$ be n mutually exclusive perfect subsets of K. Each K_i contains a subset K'_i which is similar to the linear continuum [5, p. 101].[2] Denote by D_i the subset of E_n which consists of all $\{x_k\}$ where $x_k = 0$ for $k \neq i$ and x_i assumes all real values. It follows that there exists a single-valued, monotonic increasing function g_i, with single-valued inverse, such that $g_i(D_i) = K'_i$. For convenience, write $g_i(0, 0, \cdots, x_i, \cdots, 0) = g_i(x_i)$.

Now, for any point $(x_1, \cdots, x_n)$ of E_n, let $f(x_1, \cdots, x_n) = \sum_{i=1}^{n} g_i(x_i)$. Since each g_i is single-valued and monotonic increasing in x_i, $f(x_1, \cdots, x_n)$ is single-valued and monotonic increasing in each of its variables. Also, if $f(x_1^{(a)}, \cdots, x_n^{(a)}) = f(x_1^{(b)}, \cdots, x_n^{(b)})$ for two points a and b of E_n, then $\sum_{i=1}^{n} [g_i(x_i^{(a)}) - g_i(x_i^{(b)})] = 0$. Since every finite subset of $[\psi]$ is algebraically independent, and $K'_i \cdot K'_j = 0$ for $i \neq j$, it follows that $g_i(x_i^{(a)}) = g_i(x_i^{(b)})$, $i \leq n$. But each g_i has a single-valued inverse, so that $x_i^{(a)} = x_i^{(b)}$, $i \leq n$. Therefore, f has a single-valued inverse, which completes the proof of the theorem.

5.11. Corollary. *The λ-dimension of $P_n(E_n)$ exists for every finite n.*

The proof of this follows immediately from the preceding theorem and the fact that any extension of $P'_n(E_n)$ is also an extension of $P_n(E_n)$.

We now prove the following theorem.

5.2. Theorem. *The λ-dimension of $P'_2(E_2)$ is $\aleph_0$.*

Proof. It was shown in the proof of 5.1 that one λ-extension of $P'_2(E_2)$ is of the form $f(x, y) = g(x) + h(y)$, where (x, y) is a point in E_2 and the values of g and h lie in mutually exclusive subsets of $[\psi]$. Consider now the function $f_{pq}(x, y) = pg(x) + qh(y)$, where p and q are any positive integers. The funtion f_{pq} is monotonic increasing in each of x and y, since pg and qh are, and f_{pq} is single-valued. Furthermore, by the same reasoning as in 5.1, f_{pq} has a single-valued inverse. Hence, for any pair of integers p and q, f_{pq} is a λ-extension of $P'_2(E_2)$.

We will now show that the countable totality of functions f_{pq} for all possible pairs of integers p and q realizes $P'_2(E_2)$. To see this, let (x_1, y_1), (x_2, y_2) be any two distinct points of E_2 which are non-comparable in $P'_2(E_2)$. Since $(x_1, y_1) \phi (x_2, y_2)$ either (1) $x_1 < x_2$ and $y_2 < y_1$ or (2) $x_2 < x_1$ and $y_1 < y_2$. In case (1), $[h(y_1) - h(y_2)]/[g(x_2) - g(x_1)]$ is always defined and greater than 0. Therefore, for any pair of non-comparable points (1),

[2] Every perfect set contains a subset which is similar to the linear continuum.

it is possible to choose two pairs of integers p, q and p', q' such that $p/q < [h(y_1) - h(y_2)]/[g(x_2) - g(x_1)] < p'/q'$. But $(x_2, y_2) < (x_1, y_1)$ in f_{pq}, for $pg(x_2) + qh(y_2) < pg(x_1) + qh(y_1)$, and $(x_1, y_1) < (x_2, y_2)$ in $f_{p'q'}$, for $p'g(x_1) + q'h(y_1) < p'g(x_2) + q'h(y_2)$. If (x_1, y_1), (x_2, y_2) satisfy case (2), a consideration of $[h(y_2) - h(y_1)]/[g(x_1) - g(x_2)]$ will lead to the same conclusion for pairs of non-comparable elements of this type.

Now, suppose that the λ-dimension of $P'_2(E_2)$ is finite, and equal to n. Let $f_1, \cdots, f_n$ be n λ-extensions of $P'_2(E_2)$ which realize it. If $p = (x_1, y_1)$, $q = (x_2, y_2)$ are points of E_2, and $x_1 < x_2$, then we shall say that q lies to the right of p in E_2. Let $A_k = E_p[p \varepsilon E_2$ and $p < Q$ in f_k, where Q is the set of all points q which lie to the right of p in $E_2]$. We shall call a vertical in E_2 any set of points (x, y) for which x is constant and y takes all real values. Denote the set of all verticals in E_2 by V. We remark that if $v \varepsilon V$ contains a point a of A_k, then all the points of v which precede a in $P'_2(E_2)$ also belong to A_k. Let $V_k \subset V$ be the set of verticals each of which contains at least one point of A_k. Suppose V_k to be non-denumerable, and consider on each $v \varepsilon V_k$ a pair of points belongings to A_k (each vertical of V_k will certainly contain at least two such points). This non-denumerable set of pairs of points determines, in f_k, a non-denumerable set of non-overlapping intervals; for, if p, q is a pair of points on $v \varepsilon V_k$ belonging to A_k, and p', q' is another such pair of points on $v' \varepsilon V_k$, where v' lies to the right of v in E_2 and $p > q$, $p' > q'$ in $P'_2(E_2)$, then $f_k(q) < f_k(p) < f_k(q') < f_k(p')$. This contradicts the assumption that f_k is a λ-extension of $P'_2(E_2)$, and, therefore, V_k is countable.

Now, consider $V' = V - \sum_{k=1}^{n} V_k$. If $v' \varepsilon V'$, then no point of v' belongs to $\sum_{k=1}^{n} A_k$. Let $p = (x_1, y_1)$ be a point of v'. Then there exist points q_k, $k = 1, \cdots, n$, to the right of p in E_2 such that $f_k(q_k) < f_k(p)$. Let $q = (x_2, y_2)$ be a point to the right of p such that $q < q_k$ in $P'_2(E_2)$, $k \leq n$. Then $f_k(q) < f_k(p)$, and $q < p$ in $P'_2(E_2)$. This, however, is impossible, since $x_1 < x_2$. Therefore, the λ-dimension of $P'_2(E_2)$ is not less than $\aleph_0$, and the theorem is proved.

5.21. COROLLARY. *The λ-dimension of $P'_n(E_n)$ is $\aleph_0$ for every finite n.*

Proof. That the λ-dimension of $P'_n(E_n)$ is not greater than $\aleph_0$ follows from a repetition of the first part of the argument, extended to n variables, used in 5.2. To see that λ-dim $P'_n(E_n) \geqq \aleph_0$, we need merely remark that $P'_2(E_2)$ can be mapped by a similarity transformation on to a subset of $P'_n(E_n)$; namely by the mapping which assigns to each point (x, y) of E_2 the point $(x, y, 0, 0, \cdots, 0)$ of E_n.

We now prove a corresponding theorem for $P_2(E_2)$.

5.3. THEOREM. λ-dim $P_2(E\ ) = c$, *the power of the continuum.*

Proof. Since $\bar{\bar{E}}_2 = c$, λ-dim $P_2(E_2) \leq c$. Now, suppose that λ-dim $P_2(E_2) < c$, and that $P_2(E_2)$ is realized by the set of λ-extensions $F = [f]$, where $\bar{\bar{F}} < c$. Let V be any set of verticals in E_2 of power c, and consider in E_2 the intersection of V with two horizontal lines $y = y_1$, and $y = y_2$. Let U be the set of upper points of this intersection, and L the set of lower points; and denote by u and l, where $u \varepsilon U$, $l \varepsilon L$, corresponding points on the same vertical v. Since λ-dim $P_2(E_2) < c$, there exists an $f \varepsilon F$ such that $f(u) < f(l)$ for a set of pairs u, l of power c. This, however, implies that f contains a non-denumerable set of non-overlapping intervals. For, if v' lies to the right of v, and $u, l \varepsilon v$, $u', l' \varepsilon v'$ are such that $f(u) < f(l)$ and $f(u') < f(l')$, then $f(u) < f(l) < f(u') < f(l')$. This contradicts the assumption that f is a λ-extension of $P_2(E_2)$, and, therefore, λ-dim $P_2(E_2) = c$.

5.31. COROLLARY. λ-dim $P_n(E_n) = c$ *for every finite* n.

Proof. Since $\bar{\bar{E}}_n = c$, λ-dim $P_n(E_n) \leq c$. That λ-dim $P_n(E_n) \geq c$ follows from the fact that $P_2(E_2)$ can be mapped by a similarity transformation on to a subset of $P_n(E_n)$; namely, by the mapping which assigns to each point (x, y) of E_2 the point $(x, y, x, x, \cdots, x)$ of E_n.

6. Subsets of E_n for which P_n and P'_n have finite λ-dimension. The question arises as to whether there exist subsets M_n of E_n for which the λ-dimension of $P_n(M_n)$ or $P'_n(M_n)$ is finite and greater than n. The purpose of what follows is to show that this question is to be answered in the negative. We restrict ourselves here to the case in which n is finite.

For terminological convenience, we shall, in what follows, call any set v in E_n a vertical parallel to the x_i-axis if for any two points a and b of v, $x_k^{(a)} = x_k^{(b)}$ for $k \neq i$ and $x_i^{(a)} \neq x_i^{(b)}$.

The following theorem provides a characterization of those subsets of E_n for which the λ-dimension of P_n and P'_n does not exceed n.

6.1. THEOREM. *Let M_n be any subset of E_n which, except for a denumerable set of verticals, has at most one point in common with any vertical. Then the λ-dimension of $P_n(M_n)$ or $P'_n(M_n)$ is at most n.*

Proof. We prove the theorem first for $P'_n(M_n)$. Let $\bar{N}_i$ be the projection of M_n on the x_i-axis, and let N_i be a linear order defined by means of $\bar{N}_i$ as follows. A point $\bar{u}$ of $\bar{N}_i$ which is the projection of a single point

of M_n goes over into a point u of N_i. A point $\bar{v}$ of $\bar{N}_i$ which is the projection of more than one point of M_n goes over into an interval v of N_i. The order relation of the points u and the intervals v in N_i is the same as the order relation of points $\bar{u}$ and $\bar{v}$ in $\bar{N}_i$.

It is clear that N_i, as an ordered set of points (when the points of the interals v are considered), is similar to a subset of the linear continuum.

We now transform M_n into a subset L_i of N_i as follows. Any point of M_n whose projection is a point $\bar{u}$ (in $\bar{N}_i$) goes over into the corresponding point u; all the points of M_n whose projection is the same point $\bar{v}$ go over into a similar subset of the interval v ("below"-"above" among such points of M_n goes over into "left"-"right" in this subset of v). It is clear that L_i is a λ-extension of $P'_n(M_n)$, and that the set of extensions $L_1, \cdots, L_k$ realize this partial order. Thus, if a and b are any two points of M_n such that $x_i^{(a)} < x_i^{(b)}$ and $x_j^{(a)} > x_j^{(b)}$ $(i \neq j,\ i, j \leq n)$ then the image of a precedes the image of b in L_i and succeeds it in L_j.

For the case of $P_n(M_n)$ the proof of this theorem is almost identical with the above, except that the phrase: "below"-"above" goes over into "left"-"right" is replaced by the phrase: "below"-"above" goes over into "right"-"left."

We now show that every subset of E_n which does not satisfy the conditions of Theorem 6.1 does not have finite λ-dimension.

Let A_n be any subset of E_n which has two points in common with each of a non-denumerable set of verticals V. Without loss of generality we may consider the set of verticals V to be parallel to the x_1-axis of E_n. For any two distinct points l and u of A_n which lie on the same vertical of V, and such that $x_1^{(l)} < x_1^{(u)}$, call l the lower point and u the upper point. We prove first

6.2. Lemma. *For every λ-extension f of $P'_n(A_n)$, there exists a non-denumerable subset A' of A_n, such that $A' = U' + L'$, and $f(L') < f(U')$, where U' is the set of upper and L' the set of lower points of A'.*

Proof. Let f be any λ-extension of $P'_n(A_n)$. Consider the function $g(u, l) = f(u) - f(l)$, where we denote by (u, l) a pair of points which lie on the same vertical of V. Since $u > l$ in $P'_n(A_n)$, $f(u) > f(l)$, and $g(u, l) > 0$ for every pair (u, l). Therefore, $g(u, l) > 0$ for a non-denumerable set of values.

Now, let M_i be the set of pairs (u, l) for which $g(u, l) > 1/i$. Since $\sum_{i=1}^{\infty} M_i = A_n$ is non-denumerable, there exists an M_k which is non-denumerable.

Let U_k be the set of upper points of M_k. Since $f(U_k)$ is non-denumerable, all but a denumerable set of points of $f(U_k)$ are condensation points of $f(U_k)$ both from the right and from the left. Delete this denumerable set and call the rest $f(\bar{U}_k)$. Let $f(u_1)$ be a point of $f(\bar{U}_k)$. Consider an interval $N_{f(u_1)}$ about $f(u_1)$ which is of diameter $1/2k$. Let $f(\bar{U}_k) \cdot N_{f(u_1)} = f(U')$, and let L' be the set of lower points corresponding to U'. Now $U' + L' = A'$ is non-denumerable, and we assert that $f(L') < f(U')$. Suppose this is not the case. Then there exists $f(l) \varepsilon f(L')$ such that $f(l) \varepsilon N_{f(u_1)}$. But, since $f(u)$, where u is the upper point corresponding to l, also belongs to $N_{f(u_1)}$, it follows that $f(u) - f(l) < 1/2k$. This, however, contradicts the fact that $f(u) - f(l) > 1/k$ since u, l are points of M_k.

6.3. THEOREM. λ-dim $P'_n(A_n) = \aleph_0$.

Proof. By Corollary 5.21, λ-dim $P'_n(A_n) \leq \aleph_0$. Suppose that the λ-dimension of $P'_n(A_n)$ is finite and that this partial order is realized by the λ-extensions $f_1, \cdots, f_m$. Then, by Lemma 6.2, there exists a non-denumerable subset A' of A_n such that $A' = U' + L'$, and $f_1(L') < f_1(U')$. Applying the lemma to A', there exists a non-denumerable subset A^2 of A' such that $A^2 = U^2 + L^2$, and $f_2(L^2) < f_2(U^2)$, etc. Therefore, there exists a non-denumerable subset A^m of A_n such that $A^m = U^m + L^m$, where $U^m \subset U^{m-1} \subset \cdots \subset U'$, $L^m \subset L^{m-1} \subset \cdots \subset L'$, and $f_i(L^m) < f_i(U^m)$ $(i \leq m)$. This, however, implies that $L^m < U^m$ in $P'_n(A_n)$, which is clearly impossible.

6.4. THEOREM. λ-dim $P_n(A_n) = c$, *the power of the continuum.*

Proof. By Corollary 5.31, λ-dim $P_n(A_n) \leq c$. Suppose that λ-dim $P_n(A_n) < c$, and let $P_n(A_n)$ be realized by the set F of λ-extensions $[f]$, where $\bar{\bar{F}} < c$. Then there exists an $f \varepsilon F$ such that $f(u) < f(l)$ for a non-denumerable set of pairs (u, l), each pair lying on the same vertical. Using the same argument as that in 6.2 it follows that there exists a non-denumerable subset A' of A_n such that $A' = U' + L'$, and $f(U') < f(L')$. It is easily seen that there exist u and l, u' and l', of A', such that the i-th coordinate of u' is greater than the i-th coordinate of l. This gives $f(u) < f(l) < f(u')$, which contradicts the fact that $f(U') < f(L')$.

7. The λ-dimension of $P(\aleph_0)$. It was shown in 2 that dim $P(\alpha) = \alpha$. We shall now discuss this partial order from the point of view of the λ-dimension. Clearly, $P(\alpha)$, for $\alpha > \aleph_0$, does not have λ-dimension (assuming the continuum hypothesis). Again, if α is a finite cardinal, then, by 4.22, $P(\alpha)$ has λ-dimension, and λ-dim $P(\alpha) = \dim P(\alpha)$. It remains to consider the case $\alpha = \aleph_0$.

7.1. THEOREM. *$P(\aleph_0)$ has λ-dimension, and λ-dim $P(\aleph_0) = \aleph_0$.*

Proof. Without loss of generality, we may consider $P(\aleph_0)$ to be defined on the set of subsets $\mathfrak{N}$ of the set of natural numbers D. We show first that there is a subset $M_{\aleph_0}$ of $E_{\aleph_0}$ such that $P(\aleph_0) \sim P'_{\aleph_0}(M_{\aleph_0})$. If N is any subset of D, associate with N the sequence $\{x_k\}$, where $x_k = 1$ or 0 according as $k \varepsilon N$ or $k \not\varepsilon N$. The sequences $\{x_k\}$ form a subset $M_{\aleph_0}$ of $E_{\aleph_0}$. It is obvious that the association is $1 - 1$. Let $N = \{n_i\}$, $N' = \{n'_i\}$, where n_i, n'_i are integers, be any two subsets of D, with images $\{x_k\}$, $\{x'_k\}$ in $M_{\aleph_0}$. If $\{n_i\} < \{n'_i\}$ in $P(\aleph_0)$, $N \subset N'$, and $x_k \leq x'_k$ for all k. Since N is a proper subset of N', there exists an $n'_j \not\varepsilon N$, and $x_{n'_j} < x'_{n'_j}$, so that $\{x_k\} < \{x'_k\}$ in $P'_{\aleph_0}(M_{\aleph_0})$. If, on the other hand, $\{n_k\} \phi \{n'_k\}$ in $P(\aleph_0)$, there exist $n_l \not\varepsilon N'$ and $n'_m \not\varepsilon N$, so that $\{x_k\} \phi \{x'_k\}$ in $P'_{\aleph_0}(M_{\aleph_0})$. Therefore, $P(\aleph_0) \sim P'_{\aleph_0}(M_{\aleph_0})$.

To show that $P'_{\aleph_0}(M_{\aleph_0})$ has a λ-extension, associate each point $a = \{x_k\}$ of $M_{\aleph_0}$ with the real number $r_a = \sum_{i=1}^{\infty} x_i/3^i$. Define the linear order K on $M_{\aleph_0}$ by putting $a < b$ in K if and ony if $r_a < r_b$, where a and b are any two points of $M_{\aleph_0}$. Clearly, K is a λ-extension of $P'_{\aleph_0}(M_{\aleph_0})$, for if $a < b$ in $P'_{\aleph_0}(M_{\aleph_0})$, then $x_i^{(a)} \leq x_i^{(b)}$ (all i), and $x_j^{(a)} = 0$, $x_j^{(b)} = 1$ for some j, so that $r_a < r_b$. It follows that $P(\aleph_0)$ has λ-dimension.

By 2.1, $\dim P(\aleph_0) = \aleph_0$, so that λ-dim $P(\aleph_0) \geq \aleph_0$. On the other hand, a repetition of the argument of 2.1, where now L'_x, L''_x are λ-extensions of P'_x and P''_x, will show that λ-dim $P(\aleph_0) \leq \aleph_0$, which completes the proof.

UNIVERSITY OF MICHIGAN

REFERENCES.

1. Ben Dushnik and E. W. Miller, "Partially ordered sets," *American Journal of Mathematics*, vol. 63 (1941), pp. 600-610.

2. Edward Szpilrajn, "Sur l'extension de l'ordre partiel," *Fundamenta Mathematicae*, vol. 16 (1930), pp. 386-389.

3. J. von Neumann, "Ein System algebraisch unabhängiger Zahlen," *Mathematische Annalen*, vol. 99 (1928), pp. 134-141.

4. S. Rusciewicz and W. Sierpinski, "Sur un ensemble parfait qui a avec toute sa translation au plus un point commun," *Fundamenta Mathematicae*, vol. 19 (1932), pp. 17-21.

5. Hausdorff, "*Grundzüge der Mengenlehre*," 1st ed. 1914.

SEMIGROUPS CONTAINING MINIMAL IDEALS.*

By A. H. Clifford.

The purpose of the present note is to extend to infinite semigroups S certain results of Suschkewitsch [1] concerned with what he calls the "Kerngruppe" K of a finite semigroup S. Rees [2] has pointed out that K may be described as the intersection of all two-sided ideals of S. Although Rees extends Suschkewitsch's determination of the structure of K to infinite "completely simple" K, he discusses (*loc. cit.*, p. 392) the Kerngruppe of a non-simple S only in the case S finite.

We replace finiteness by the condition that S contain at least one minimal left (or right) ideal; K is then a simple subsemigroup of S without zero. If S contains both minimal left and right ideals, K is completely simple in the sense of Rees. If S contains exactly one minimal left ideal and exactly one minimal right ideal, then K is the group of "zeroids" [3] of S.

1. The kernel (Kerngruppe) of a semigroup. By a semigroup we mean a system S of elements $a, b, c, \cdots$ closed under an associative binary operation: $(ab)c = a(bc)$. By a left ideal of S we mean a non-vacuous subset A of S with the property $SA \subseteq A$, i. e. $sa \in A$ for every $s \in S$, $a \in A$. A left ideal is minimal if it does not contain any proper left subideal of S. Similar definitions hold for "right" instead of "left." A (two-sided) ideal is a subset of S which is both a left and a right ideal.

A semigroup S is said to be simple if it contains no proper two-sided ideal, except possibly one consisting of a single element z. Such an element z has the property $az = za = z$ for all a in S, and is termed a zero element. If S contains no proper two-sided ideal at all, it is called a "simple semigroup without zero."

If the intersection K of all the two-sided ideals of a semigroup S is not vacuous, we shall call K the *kernel* (Suschkewitsch's Kerngruppe) of S.

* Received October 3, 1947.

[1] A. Suschkewitsch, "Über die endlichen Gruppen ohne das Gesetz der eindeutige Umkehrbarkeit," *Mathematische Annalen*, vol. 99 (1928), pp. 30-50.

[2] D. Rees, "On semi-groups," *Proceedings of the Cambridge Philosophical Society*, vol. 36 (1940), pp. 387-400.

[3] A. H. Clifford and D. D. Miller, "Semigroups having zeroid elements," *American Journal of Mathematics*, vol. 70 (1948), pp. 117-125.

If we say that a semigroup "has a kernel" we shall mean that K is not vacuous. An example of a semigroup not having a kernel is that of the infinite cyclic semigroup (the natural numbers under addition). K is of course itself an ideal, and is contained in every ideal of S. (We drop the modifier "two-sided" for the rest of this section.)

If a semigroup S contains a minimal ideal M then M is the kernel of S. For if A is any ideal of S, the intersection $A \cap M$ is non-vacuous, since it contains the product AM. Since M is minimal, $A \cap M = M$, i. e. $M \subseteq A$.

THEOREM 1.1. *If a semigroup S has a kernel K, then K is a simple subsemigroup of S without zero.*

Proof. Since K is an ideal it is *a fortiori* a subsemigroup of S. To show that it is simple without zero, let A be any ideal of K. KAK is an ideal of S contained in K. From the minimality of K, $KAK = K$. But $KAK \subseteq A$, whence $K = A$, i. e. A is not a proper ideal of the semigroup K.

K may consist of a single element, which is the case if and only if S has a zero element.

2. Semigroups having minimal left ideals.

LEMMA 2.1. *If L is a minimal left ideal of a semigroup S, and c is any element of S, then Lc is also a minimal left ideal of S.*

Proof. Lc is clearly a left ideal of S. Suppose that M is a left subideal of Lc. Let L_1 be the set of all elements a_1 of L such that $a_1c \in M$. If $s \in S$, $sa_1c \in M$, so that $sa_1 \in L_1$. Hence L_1 is a left ideal. Since L is minimal, $L_1 = L$. Hence $Lc = M$, and Lc has no proper subideal.

LEMMA 2.2. *A two-sided ideal A of S contains every minimal left ideal L.*

Proof. AL is a left ideal of S contained in L. Since L is minimal, $AL = L$. But $AL \subseteq A$, so that $L \subseteq A$.

THEOREM 2.1. *If a semigroup S contains at least one minimal left ideal, then it has a kernel K. K is the class sum of all the minimal left ideals of S.*

Proof. Let M be the class sum of all the minimal left ideals of S. M is non-vacuous by hypothesis, and is clearly a left ideal of S. To show that it is also a right ideal, let $c \in S$ and $m \in M$. By definition of M, m belongs

to a minimal left ideal L of S, and $mc \in Lc$. By Lemma 2.1, Lc is also a minimal left ideal of S, and hence is contained in M. Thus $mc \in M$.

By Lemma 2.2, M is contained in every two-sided ideal of S, and since it is one itself, it must be the kernel of S.

THEOREM 2.2. *Under the hypothesis of Theorem* 2.1, *every left ideal of K is also a left ideal of S.*

Proof. Let A be a left ideal of K. Each element a of A belongs to some minimal left ideal L of S. Ka is a left ideal of S contained in L, whence $Ka = L$. In particular, $a \in Ka$. This implies $A \subseteq KA$, hence $A = KA$. But KA is a left ideal of S.

It is clear from this theorem that every minimal left ideal of S is also a minimal left ideal of K, and *vice versa.*

THEOREM 2.3. *Under the hypothesis of Theorem* 2.1, *every left ideal of S contains at least one minimal left ideal of S.*

Proof. Let A be a left ideal of S. KA is a left ideal of S contained in K, and hence consisting of the class sum of certain of the minimal left ideals L of S. But KA is also contained in A. Hence A contains at least one L.

A semigroup S will be called *left simple* if it contains no proper left ideal except possibly that consisting of the zero element only. If the latter possibility is also excluded, S is a "left simple semigroup without zero." This is the case if and only if the equation $ya = b$ is always solvable in y, for any given pair of elements a, b of S.

We remark that a semigroup S which is both left and right simple and without zero is a group.[4]

THEOREM 2.4. *A minimal left ideal L of a semigroup S is a left simple subsemigroup of S without zero.*

Proof. Let a and b be elements of L. La is a left ideal of S contained in L, whence $La = L$ by the minimality of L. This implies $b \in La$, i. e., the equation $ya = b$ is solvable for y in L.

If S is a finite left simple semigroup without zero, then the solution of $ya = b$ is unique, and conversely uniqueness (if $y_1 a = y_2 a$ then $y_1 = y_2$) implies existence. Such an S is called a "Linksgruppe" by Suschkewitsch.

---

[4] E. V. Huntington, "Simplified definition of a group," *Bulletin of the American Mathematical Society,* vol. 8 (1901-02), pp. 296-300.

For infinite S, it is of some interest to note that a left simple semigroup S without zero is a left group (i. e., $ya = b$ is uniquely solvable) if and only if S contains an idempotent element. The structure of such semigroups was determined by the author,[5] unaware of the earlier work of Suschkewitsch (which, however, applied only to finite semigroups).

3. Semigroups containing both minimal left and right ideals. Needless to say, the results of the preceding section have obvious left-right duals, applicable to semigroups containing minimal right ideals. The following theorem is an immediate consequence of Theorem 2.1 and its dual, together with a repetition of Theorem 1.1.

THEOREM 3.1. *If a semigroup S contains at least one minimal left ideal and at least one minimal right ideal, then the class sum of all the minimal left ideals of S coincides with that of all its minimal right ideals, and constitutes the kernel K of S. K is a simple subsemigroup of S without zero.*

The purpose of this section is to show that K is *completely* simple in the sense of Rees, and consequently of known structure. The proof is a slight modification of one suggested by R. H. Bruck for **2** of the paper cited in footnote 3.

An idempotent $f\,(f^2 = f)$ is said to be *under* another one e if $ef = fe = f$. An idempotent e is *primitive* if there is no non-zero idempotent under e. A simple semigroup S is said to be *completely simple* if (1) every idempotent element of S is primitive, and (2) for each $a \varepsilon S$ there exist idempotents e and f in S such that $ea = af = a$. Rees shows that every finite simple semigroup is completely simple.

In the following lemmas, R denotes a minimal right ideal and L a minimal left ideal of a semigroup S. The product RL is clearly a non-vacuous subset of the intersection $R \cap L$.

LEMMA 3.1(R). *If $a \varepsilon R$ and $b \varepsilon RL$, the equation $ax = b$ has a solution x in RL.*

Proof. aR is a right ideal of S contained in R, so that $aR = R$ by the minimality of R. Hence $aRL = RL$.

LEMMA 3.1(L). *If $a \varepsilon L$ and $b \varepsilon RL$, the equation $ya = b$ has a solution y in RL.*

---

[5] A. H. Clifford, "A system arising from a weakened set of group postulates," *Annals of Mathematics*, vol. 34 (1933), pp. 865-871.

Proof. La is a left ideal $\subseteq L$, whence $La = L$ and $RLa = RL$.

LEMMA 3.2. *RL is a group.*

Proof. $RL \cdot RL \subseteq RL \cdot L \subseteq RL$, whence RL is closed. Let a and b be any two elements of RL. Since a belongs to both R and L, we conclude from Lemmas 3.1(R) and (L) that the equations $ax = b$ and $ya = b$ are both solvable for x and y in RL. Hence RL is a group.[4]

LEMMA 3.3. *Let e be the identity element of the group RL. Then $R = eS$, $L = Se$, and $R \cap L = eSe$.*

Proof. Since $e \in R$, eS is a right ideal contained in R, hence equal to R by the minimality of R. Similarly, $Se = L$. Since R consists of all x in S such that $ex = x$, and L of all y such that $ye = y$, $R \cap L$ consists of all w in S such that $ew = we = e$, that is, $R \cap L = eSe$.

LEMMA 3.4. *$R \cap L = RL$.*

Proof. We need of course show only that $R \cap L \subseteq RL$. Let $a \in R \cap L$, and let e be the identity element of the group RL. By Lemma 3.1(R) we can solve $ax = e$ for x in RL. If x^{-1} is the inverse of x in RL, we have $ae = axx^{-1} = ex^{-1} = x^{-1}$. But $a \in eSe$ by Lemma 3.3, so that $ae = a$. Hence $a(= x^{-1}) \in RL$.

LEMMA 3.5. *The identity element e of the group RL is a primitive idempotent.*

Proof. Suppose f were an idempotent under e, so that $ef = fe = f$. By Lemma 3.3, $f \in R \cap L$. By Lemma 3.4, $f \in RL$, and since a group can contain only one idempotent, $f = e$.

THEOREM 3.2. *Under the hypothesis of Theorem 3.1, the kernel K of S is a completely simple semigroup without zero.*

Proof. Each element of K belongs to exactly one minimal left ideal L (since the intersection of two such is vacuous) and to exactly one minimal right ideal R, hence to exactly one of the groups $R \cap L$. K is thus the class sum of the non-overlapping groups $R \cap L$. Each idempotent element of K must belong to one of these groups, and so must be the identity element thereof. By Lemma 3.5, condition (1) for complete simplicity is established.

Each element a of K belongs to one of the groups $R \cap L$. Let e be the identity element thereof. Then $ea = ae = a$. Hence condition (2) holds (with $f = e$).

4. Semigroups containing exactly one minimal left and/or right ideal; zeroid elements. In this concluding section we tie in the foregoing results with those of the paper cited in footnote 3, which we shall refer to as *SHZE*. As defined there, an element l of S is a *left zeroid* of S if, for any $a \varepsilon S$, the equation $xa = l$ has a solution x in S. It is easily shown (**1** of *SHZE*) that S contains a left zeroid if and only if it contains a "universally minimal" left ideal L, i. e. one contained in every left ideal of S, and that L consists of all the left zeroids of S. Now it is an immediate corollary of Theorem 2. 3 that if S contains exactly one minimal left ideal L, then L is universally minimal. From these remarks, together with Theorems 2. 1 and 2. 4, we conclude:

THEOREM 4. 1. *A semigroup S contains a left zeroid element if and only if it contains exactly one minimal left ideal L. L consists of all the left zeroids of S, and is contained in every left ideal of S. L is a right as well as left ideal, being in fact the kernel of S, and is a left simple subsemigroup of S without zero.*

Combining Theorem 4. 1 with its left-right dual, and using the remark preceding Theorem 2. 4, we obtain:

THEOREM 4. 2. *If a semigroup S contains exactly one minimal left ideal L and exactly one minimal right ideal R, then the two coincide, and constitute the kernel K of S. K is a group, and consists of all the (left and right) zeroid elements of S.*

This differs from Theorems 1 and 2 of **2** of *SHZE* only in its apparently weaker hypothesis, that L and R are "locally" rather than "universally" minimal. It should be remarked, however, that the proof of Theorem 4. 2 makes no use of Theorem 3. 2, although the result is a consequence thereof.

THE JOHNS HOPKINS UNIVERSITY.

FREE PRODUCTS OF GROUPS.*

By B. L. van der Waerden.

Some twenty years ago, I discussed with Artin and the late Schreier the existence of Free Groups and Free Products of Groups, and the difficulty of establishing the associative law in them. I suggested that it might be easier to define the elements of the free product not as words but as permutations of words, because the associative law is trivial for permutations. We agreed that this idea was correct, and that, if fully worked out, it would be substantially equivalent to the classical proof, due to Artin and Schreier.[1] After this, none of us thought of it any more.

Quite recently, Artin published a new proof of the existence of free products.[2] In his proof, the associative law is trivial as well as in mine, but the main difficulties arise from the discussion of equality. Now I should like to publish my proof, in which neither the associative law nor the definition of equality causes any difficulty.

Let a set of groups H be given, and let $a, b, \ldots$ denote their elements. From these letters, *words* may be formed, with the convention that no two consecutive letters shall belong to the same group H, and that no letter be the unity element. A word may be empty, i. e. contain no letters at all.

To any letter a (or $b, \ldots$) let there be assigned an operation A (or $B, \ldots$), operating on words and transforming them into words, as follows: If a word begins with a letter b from the same group H to which a belongs, the operation A consists in replacing b by the letter $c = ab$, omitting it if c is the unity element. If the word is empty or begins with a letter from a different group H, the operation A consists in prefixing the letter a.

If a and b are letters from the same group H and if $ab = c$, then $AB = C$. In other words, the operation B and A successively applied to any word, give the same effect as C. If the word is empty or begins with a letter from a different group H, this is obvious. If it begins with a letter d from the same group H, the operation AB replaces d by $a(bd) = (ab)d = cd$, which is just what C does.

* Received November 4, 1947.

[1] O. Schreier, *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, vol. 5 (1927), pp. 164-168.

[2] E. Artin, *American Journal of Mathematics*, vol. 69 (1947), pp. 1-4.

As a special case, we have: If a is the inverse of b, A is the inverse of B. Thus, every operation A has an inverse operation A^{-1}. Hence the operations A are permutations of words.

These permutations generate a group of permutations. The elements of the group are finite products $ABC\cdots$. If two factors arising from the same group H are consecutive, they may be replaced by one factor. So any product $ABC\cdots$ may be written in a *reduced form,* in which no two consecutive factors belong to the same group and no factor is unity. Now we can prove the following statement:

If two reduced products $ABC\cdots$ and $A'B'C'\cdots$ are equal, they must consist of exactly the same factors in the same order.

Proof. Applying the operations $ABC\cdots$ and $A'B'C'\cdots$ to the empty word, we obtain the words $abc\cdots$ and $a'b'c'\cdots$. These two words must be equal, hence $a=a'$, $b=b'$, etc.

Thus the group G of the reduced products $ABC\cdots$ has all the properties required of the free product of the groups H, viz.:

1) G contains, corresponding to any group H, a subgroup isomorphic to H;

2) Every element of G can be written as a reduced product of elements of these subgroups, in which no two consecutive factors belong to the same subgroup;

3) Two reduced products are equal only if they consist of the same factors in the same order.

THE JOHNS HOPKINS UNIVERSITY.

ON NON-CONSERVATIVE LINEAR OSCILLATORS OF LOW FREQUENCY.*

By PHILIP HARTMAN and AUREL WINTNER.

1. In the differential equation

$$x'' + f(t)x = 0, \tag{1}$$

let $f = f(t)$, where $0 \leqq t < \infty$, be a positive, non-increasing, continuous function,

$$f > 0 \text{ and } df \leqq 0, \tag{2}$$

and let

$$f(t) \to 0 \text{ as } t \to \infty. \tag{3}$$

By a solution of (1) will be meant a real-valued solution $x = x(t)$ which is not identically zero. A solution of (1) will be called non-oscillatory or oscillatory according as it possesses a finite or an infinite set of zeros; correspondingly, the differential equation (1) will be called non-oscillatory or oscillatory according as one (hence, every) solution is non-oscillatory or oscillatory.

The object of this paper is a discussion of the boundedness of the solutions of (1). Subject to exceptions, there are two general patterns. First, if $f(t)$ is "very small," then the solutions of (1) are non-oscillatory and behave like the solutions of $x'' = 0$; namely, some solutions are bounded and some are unbounded. Secondly, if the decrease of $f(t)$ to 0 is sufficiently smooth and so slow as to make the solutions of (1) oscillatory, then all solutions of (1) are unbounded. The following theorem describes the situation more fully:

Let $f(t)$, where $0 \leqq t < \infty$, be a continuous function satisfying (2) *and* (3). *Then*

(i) *at least one solution $x = x(t)$ of* (1) *is unbounded;*

(ii) *there exists a non-trivial* ($\not\equiv 0$), *non-oscillatory, bounded solution of* (1) *if and only if*

$$\int_0^\infty t f(t)\,dt < \infty; \tag{4}$$

* Received February 10, 1948.

in which case (1) *possesses a pair of solutions,* $x = x_1(t)$ *and* $x = x_2(t)$, *satisfying the asymptotic relations*

$$(5_1) \qquad x_1(t) \sim 1 \text{ and } x'_1(t) = o(1)$$

and

$$(5_2) \qquad x_2(t) \sim t \text{ and } x'_2(t) \sim 1$$

as $t \to \infty$;

(iii) *the differential equation can be oscillatory and still such as to possess a bounded solution* $(\not\equiv 0)$;

(iv) *all solutions of* (1) *are oscillatory and unbounded when* $f(t)$ *possesses continuous first and second derivatives satisfying*

$$(6) \qquad \limsup_{t\to\infty} -f'/f^{3/2} < 4$$

and

$$(7) \qquad (-f'/f)' \leqq 0,$$

so that the graph of

$$(8) \qquad -\log f(t) \text{ is concave downwards;}$$

(iv bis) *if* (7) *in* (iv) *is retained but* (6) *is relaxed to*

$$(6 \text{ bis}) \qquad \limsup_{t\to\infty} -f'/f^{3/2} = 4,$$

then (1) *need not be oscillatory*;

(v) *all solutions of* (1) *are unbounded if* (4) *is violated and*

$-\log f(t)$ is of regular growth.[1]

2. In the proofs of these assertions, the central rôle will be played by the "conjugate energy,"

$$(9) \qquad X = X(t) = x^2(t) + x'^2(t)/f(t),$$

[1] The term "regular growth" is meant in the sense defined by Armellini [1], as follows: A non-decreasing function $F = F(t)$, where $0 \leqq t < \infty$, is said to be of "irregular growth" if, for every $\epsilon > 0$, there exists an unbounded sequence of numbers

$$0 < a_1 < b_1 < a_2 < b_2 < \cdots$$

satisfying

$$\sum_{n=1}^{\infty} (F(b_n) - F(a_n)) < \infty \quad \text{and} \quad \limsup_{m\to\infty} \sum_{n=1}^{m-1} (a_{n+1} - b_n)/a_m < \epsilon.$$

"Regular growth" is defined as the negation of "irregular growth."

associated with any solution $x = x(t)$ of (1). The differential equation (1) implies that

(10) $$dX(t) = x'^2(t)\,d(1/f(t));$$

if $f(t)$ is not differentiable, the differentials in (10) are to be interpreted in the sense of Stieltjes integrations. In view of (10) and the second inequality in (2),

(11) $$dX \geqq 0.$$

Consequently, if the solutions of (1) are oscillatory, (9) and (11) show that the sequence of amplitudes of the half-waves of the graph of $x = x(t)$ is non-decreasing; that is, the sequence of values of $|x(t)|$, corresponding to the points where $x'(t) = 0$, is non-decreasing.

This fact, when combined with a known refinement (cf. [11], p. 518) of the standard wording of Sturm's comparison theorems, leads to the following lemma:

(vi) *Let $f(t)$ be a continuous function satisfying (2). Suppose that (1) has a solution $x = x(t) \not\equiv 0$ possessing three consecutive zeros, say $t_1 < t_3 < t_5$. Let t_2 and t_4, where $t_1 < t_2 < t_3 < t_4 < t_5$, denote the unique t-values at which $x' = 0$. Then the lengths of the t-intervals corresponding to four "quarter-waves" satisfy*

$$t_{j+1} - t_j \leqq t_{j+2} - t_{j+1}, \qquad (j = 1, 2, 3),$$

and, after the reflections and translations placing the abscissae of the maximum ordinates in coincidence, the graph of $x = |x(t)|$ for a quarter-wave $t_j \leqq t \leqq t_{j+1}$ lies under the graph of $x = |x(t)|$ for the next quarter-wave $t_{j+1} \leqq t \leqq t_{j+2}$; in other words,

$$|x(t)| \leqq |x(2t_2 - t)| \text{ for } t_1 \leqq t \leqq t_2$$

and

$$|x(t)| \leqq |x(t_2 + t_4 - t)| \text{ for } t_2 \leqq t \leqq t_3.$$

The proof of (vi) will be given for the case of the intervals (t_2, t_3) and (t_3, t_4), the other cases being similar. Suppose $x(t) > 0$ for $t_2 < t < t_3$; so that $x(t) < 0$ for $t_3 < t < t_4$. The statements to be proved are equivalent to

(12) $$y(t) \geqq x(t) \text{ for } t_2 \leqq t \leqq t_3,$$

where

$$y(t) = -x(t_4 + t_2 - t).$$

In order to prove (12), notice that

$$y'(t_2) = x'(t_4) = 0 \text{ and } x'(t_2) = 0, \tag{13}$$

while, in view of the remarks following (11),

$$y(t_2) = -x(t_4) \geqq x(t_2). \tag{14}$$

Its definition shows that $y = y(t)$ is a solution of the differential equation

$$y'' + f(t_4 - t_2 - t)y = 0. \tag{15}$$

Since it is assumed that $f(t)$ is non-increasing, the coefficient functions of (1) and (15) satisfy

$$f(t) \geqq f(t_4 - t_2 - t) \tag{16}$$

on the interval $t_2 \leqq t \leqq t^*$, where $t = t^*$ is the least t-value ($> t_2$) at which either $x(t)$ or $y(t)$ vanishes. Since $x(t)$ and $y(t)$ are non-negative, it follows from (16) in the usual manner that

$$[xy' - x'y]_{t_2}^{t} \geqq 0 \text{ for } t_2 \leqq t \leqq t^*.$$

In view of (13), the contribution of $t = t_2$ to the last bracket is 0, and so

$$xy' - x'y \geqq 0 \text{ for } t_2 \leqq t \leqq t^*.$$

Since this means that the ratio y/x possesses a non-negative derivative for $t_2 \leqq t < t^*$, it now follows from (14) that

$$y(t) \geqq (y(t_2)/x(t_2))x(t) \geqq x(t)$$

for $t_2 \leqq t \leqq t^*$. Hence, the least t-value ($> t_2$) at which $x(t)$ vanishes is not less than the corresponding t-value belonging to $y(t)$. Consequently, $t^* = t_3$. This proves (12).

3. Proof of (i). Let $x = x(t)$ and $x = z(t)$ denote a pair of linearly independent solutions of (1). Their Wronskian is a non-vanishing constant,

$$x'z - xz' = c \neq 0. \tag{17}$$

If (17) is divided by f^3 and Schwarz's inequality is applied, it follows from the definition (9) that

$$c^2/f(t) \leqq X(t)Z(t).$$

Since $X(t)$ and $Z(t)$ are non-decreasing, (3) now shows that at least one of the two functions $X(t)$, $Z(t)$ tends monotonously to infinity; say

$$X(t) \to \infty \text{ as } t \to \infty. \tag{18}$$

If the solutions of (1) are oscillatory, the proof of (i) is complete; in fact, (9) and (18) show that $|x(t)| \to \infty$ if t tends to ∞ through the values for which $x'(t) = 0$. On the other hand, if the solutions of (1) are non-oscillatory, (i) is a consequence of (ii).

4. Proof of (ii). Whether or not the continuous function $f(t)$ satisfies (2) and (3), the limitation

$$\int_0^\infty t\,|f(t)|\,dt < \infty$$

is sufficient for the existence of solutions $x = x_1(t)$ and $x = x_2(t)$ of (1) satisfying (5_1)-(5_2). As pointed out in [13], p. 66, this fact can be deduced from the results of Bôcher [3], § 4, by a suitable change of variables; for a simple proof avoiding successive approximations, cf. [13] and [14]. Hence, in order to prove (ii), it is sufficient to verify that the existence of a bounded, non-oscillatory solution $x = x(t)$ of (1) implies (4).

It can be supposed that $x(t) > 0$ for large t, say for $t \geqq t_0$, since otherwise $x(t)$ can be replaced by $-x(t)$. Then $x''(t) \leqq 0$ for $t \geqq t_0$, by (1). Hence, $x'(t) \geqq 0$ for $t \geqq t_0$, since otherwise $x(t)$ would possess a zero for $t > t_0$. Consequently, $x(t)$ tends non-decreasingly to a positive constant, while $x'(t)$ tends to 0, as $t \to \infty$.

On the other hand, from (1),

$$x(t) = x(t_0) + \int_{t_0}^{t} \left(\int_s^\infty f(r)x(r)\,dr \right) ds.$$

Since $x(t)$ tends to a positive limit and since $f(r) \geqq 0$, it is now seen that

$$\int_{t_0}^\infty \left(\int_s^\infty f(r)\,dr \right) ds < \infty.$$

On inverting the order of integration, which is permissible since $f(r) \geqq 0$, the relation (4) follows. This completes the proof of (ii).

5. Proof of (iii). The proof will consist of constructing a function $f = f(t)$ satisfying (2) and (3) but having the property that (1) is oscillatory and possesses a bounded solution. For the sake of shortness, the function to be constructed will be a non-decreasing step-function, having a sequence of isolated discontinuity points which tend to ∞. Such a function can be modified so as to be continuous and to retain the desired properties.

Let

$$a_1 > a_2 > \cdots \tag{19}$$

be a sequence of numbers satisfying

$$a_n \to 0 \text{ as } n \to \infty. \tag{20}$$

Define a function $f = f(t)$ by placing

$$f(t) = a_n^2 \text{ for } b_{n-1} \leqq t < b_n, \tag{21}$$

where

$$b_n = 2\pi \sum_{k=1}^{n} (1/a_k) \text{ for } n = 0, 1, \cdots. \tag{22}$$

If the vacuous sum denotes zero, (19)-(22) define, for $0 \leqq t < \infty$, a function $f(t)$ satisfying (2) and (3). Let

$$x(t) = \cos a_n(t - b_n) \text{ for } b_{n-1} \leqq t < b_n; \tag{23}$$

so that $x(t)$ is also defined for $0 \leqq t < \infty$. Clearly, $x(t)$ possesses a continuous first derivative for $0 \leqq t < \infty$ and a continuous second derivative on each of the intervals $b_{n-1} < t < b_n$. It is also clear that (1) is satisfied if $t \neq b_n$, where $n = 0, 1, \cdots$. Finally, the function (22) is bounded and vanishes at $t = b_{n-1} + \pi/2a_n$ and at $t = b_{n-1} + 3\pi/2a_n$. This completes the proof of (iii).

6. Proof of (iv) and (iv bis). In view of (6), there exists a constant c satisfying

$$0 < c < 4 \tag{24}$$

and

$$-f'/f^{3/2} < c \text{ for large } t. \tag{25}$$

An integration of (25) shows that

$$2/f^{\frac{1}{2}} < ct \text{ for large } t;$$

hence,

$$\liminf_{t\to\infty} ft^2 \geqq 4/c^2 > 1/4,$$

by (24). As shown by Kneser [8], this implies that (1) is oscillatory.

If $f(t)$ is defined, for $1 \leqq t < \infty$, by

$$f(t) = 1/4t^2, \tag{26}$$

then

$$-f'/f^{3/2} = 4 \text{ and } (-f'/f)' = -2/t^2 \leqq 0;$$

so that (6 bis) and (7) are satisfied. On the other hand, (1) is non-oscillatory in the case (26). This proves (iv bis).

In order to complete the proof of (iv), it remains to show that all solutions of (1) are unbounded if (6) and (7) are satisfied. To this end, introduce the new independent variable

$$s = s(t) = \int_0^t f^{\frac{1}{2}}(r)\,dr. \tag{27}$$

Then (1) becomes

$$\ddot{x} + (f'/f^{3/2})\dot{x} + x = 0, \qquad (\cdot = d/ds). \tag{28}$$

This equation is readily verified to be equivalent to the linear system

$$\dot{x} = a_{11}x + a_{12}y, \qquad \dot{y} = a_{21}x + a_{22}y, \tag{29}$$

where the coefficients $a_{ik} = a_{ik}(s)$ are defined by

$$a_{11} = -f'/2f^{3/2}, \qquad a_{12} = 1$$

and

$$a_{21} = \tfrac{1}{2}(f''/f^2 - f'^2/f^3) - 1, \qquad a_{22} = -f'/2f^{3/2}.$$

It is seen from the definition of the matrix (a_{ik}) that the characteristic numbers of the matrix $(a_{ik} + a_{ki})$ are

$$-f'/f^{3/2} \pm \tfrac{1}{2}\,|\,f''/f^2 - f'^2/f^3\,|.$$

It follows, therefore, by the argument applied in [12], p. 558, that if the smaller of these roots satisfies

$$(30) \qquad \int^{s} (-f'/f^{3/2} - \tfrac{1}{2}\,|\,f''/f^2 - f'^2/f^3\,|\,)\,ds \to \infty \text{ as } s \to \infty,$$

then

$$(31) \qquad x^2 + y^2 \to \infty \text{ as } s \to \infty$$

holds for every non-trivial solution of (29).

It will first be shown that x is unbounded by virtue of (31). Suppose the contrary. Then (6) and the definition of a_{11} imply that $a_{11}x$ is bounded as $s \to \infty$. Hence, the definition of a_{12}, the relation (31) and the first equation in (29) imply that $\dot{x}$ is unbounded. But

$$\dot{x} = x'/f^{\frac{1}{2}},$$

by (27). Consequently, the "conjugate energy" (9) is an unbounded function of s. Since (1) is oscillatory, it follows by the remarks following (11) that x is unbounded. This contradiction shows that x is unbounded by virtue of (31).

It follows that in order to complete the proof of (iv), it is sufficient to verify (30). By (27), the relation (30) becomes

$$(32) \qquad \int^{t} (-f'/f - \tfrac{1}{2}\,|\,f''/f^{3/2} - f'^2/f^{5/2}\,|\,)\,dt \to \infty \text{ as } t \to \infty.$$

Since

$$f''/f^{3/2} - f'^2/f^{5/2} = (f'/f)'/f^{\frac{1}{2}} \geqq 0$$

in virtue of (7), the absolute value signs in (32) can be omitted. Consequently, an integration by parts shows that the integral in (32) is equal to

$$-\tfrac{1}{2}f'/f^{3/2} + \int^{t} (-f'/f - f'^2/4f^{5/2})\,dt.$$

By (25) and (24), this expression is minorized by

$$\int^{t} (-f'/f)(1 - c/4)\,dt.$$

Hence, (32) follows from (24) and (3). This completes the proof of (iv).

7. It is easy to see that (iv) has the following variant:

(iv*) *The assertion* (iv) *remains true if* (6) *and* (7) *are replaced by*

(6*) $$f' = o(f^{3/2}) \text{ as } t \to \infty$$

and

(7*) $$f'' \geqq 0,$$

respectively.

Proof of (iv*). Assumption (6*) implies (6) and, as shown in the last section, (6) implies that (1) is oscillatory. Thus, the proof of (iv) shows that (iv*) will be proved if it is verified that (6*) and (7*) imply (32). The integral in (32) is minorized by

$$\int^t (-f'/f - \tfrac{1}{2}f''/f^{3/2} - \tfrac{1}{2}f'^2/f^{5/2})\,dt,$$

in view of (7*). But an integration by parts shows that the last integral equals

$$-\tfrac{1}{2}f'/f^{3/2} + \int^t (-f'/f - 5f'^2/4f^{5/2})\,dt,$$

or, by (6*),

$$o(1) + \int^t (-f'/f)(1 - o(1))\,dt.$$

Hence, (32) follows from (3). This completes the proof of (iv*). (Actually, the inequality

$$\limsup_{t\to\infty} -f'/f^{3/2} < 4/5$$

can replace the assumption (6*) in this proof of (iv*).)

8. Proof of (v). Since (v) assumes that (4) is violated, it follows from (ii) that all solutions of (1) are unbounded if (1) is non-oscillatory. Hence, it is sufficient to consider the case in which (1) is oscillatory. The proof then becomes similar to the proof of a known dual of (v), a proof proposed by Armellini [1] and completed by Tonelli [10] (the nature of such "duals" will be explained below). The proof of (v) will, therefore, be omitted.

9. Remarks. The assertions (i), (iii), and (v) have known duals. In fact, (i), (iii) and (v) remain true if the assumptions (2) and (3) are replaced by

(2*) $$f > 0 \text{ and } df \geqq 0$$

and

$$f(t) \to \infty \text{ as } t \to \infty, \tag{3*}$$

respectively, and "unbounded solution" is replaced by "solution tending to 0 as $t \to \infty$" in the assertions. In this sense, the dual of (i) was proved by Milloux [9], p. 49, in answer to a question raised by Biernacki; for a simple proof, cf. [4]. The dual of (iii) was also proved by Milloux [9], p. 50; the above proof of (iii) is based on the device used by Milloux. Finally, the dual of (v) is that referred to in the proof of (v).

Since (3*), (6*) and (7*) imply that

$$f(t + f(t)^{-\frac{1}{2}})/f(t) \to 1 \text{ as } t \to \infty,$$

(cf. the method used in [7], § 5), it follows from a theorem of Biernacki [2], p. 170, that the dual of (iv*) is true. [There can of course be no analogue of the theorem of Biernacki in which it is assumed that (2*), (3*) and $f'' \leqq 0$ hold, the latter condition being incompatible with (3).]

Assumption (6*) of (iv bis) also occurs in other questions concerning asymptotic properties of solutions of equations of the type (1); cf. [5], [6], [7].

The corresponding analogue of (vi), when combined with the "alternating series test," shows that

(vii) *if $f(t)$, where $0 \leqq t < \infty$, is a continuous function satisfying (2*) and (3*), then*

$$\int_0^\infty x(t)\,dt$$

is convergent for every solution $x = x(t)$ of (1).

The convergence of this integral is meant in the sense that

$$\lim_{T\to\infty} \int_0^T x(t)\,dt$$

exists (as a finite limit).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. Armellini, "Sopra un'equazione differenziale della dinamica," *Atti della Reale Accademia dei Lincei, Rendiconti*, ser. 6, vol. 21 (1935), pp. 111-116.

[2] M. Biernacki, "Sur l'équation différentielle $x'' + A(t)x = 0$," *Prace Matematyczno-Fizyczne*, vol. 40 (1933), pp. 163-171.

[3] M. Bôcher, "On regular singular points of linear differential equations of the second order whose coefficients are not necessarily analytic," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 40-52.

[4] P. Hartman, "On a theorem of Milloux," *American Journal of Mathematics*, vol. 70 (1948), pp. ———.

[5] ——— and A. Wintner, "The asymptotic arcus variation of solutions of linear differential equations of the second order," *American Journal of Mathematics*, vol. 70 (1948), pp. 1-11.

[6] ——— and A. Wintner, "Criteria of non-degeneracy for the wave equation," *American Journal of Mathematics*, vol. 70 (1948), pp. ———.

[7] ——— and A. Wintner, "On the asymptotic problems of the zeros in wave mechanics," *American Journal of Mathematics*, vol. 70 (1948), pp. 461-480.

[8] A. Kneser, "Untersuchungen über die reellen Nullstellen der Integrale linearer Differentialgleichungssysteme," *Mathematische Annalen*, vol. 42 (1893), pp. 409-435.

[9] H. Milloux, "Sur l'équation différentielle $x'' + A(t)x = 0$," *Prace Matematyczno-Fizyczne*, vol. 41 (1934), pp. 39-53.

[10] L. Tonelli, "Estratto di lettera al Prof. Giovanni Sansone," *Scritti Matematici offerti a Luigi Berzolari*, Padua (1936), pp. 404-405.

[11] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge (1932).

[12] A. Wintner, "Asymptotic integration constants," *American Journal of Mathematics*, vol. 68 (1946), pp. 553-559.

[13] ———, "Asymptotic integration of the adiabatic oscillator in its hyperbolic range," *Duke Mathematical Journal*, vol. 15 (1948), pp. 55-67.

[14] ———, "A norm criterion for non-oscillatory differential equations," *Quarterly of Applied Mathematics*, vol. 6 (1948), pp. ———.

THE RADICAL OF AN ALTERNATIVE ALGEBRA.* [1]

By Roy Dubisch and Sam Perlis.

Of the several equivalent definitions of the radical of an associative algebra, only one has been applied to the alternative case. This is the definition in terms of properly nilpotent quantities used by Zorn [2] in his treatment of hypercomplex rings which yields an ideal $\mathfrak{N}$. In this paper we consider also the sum, $\mathfrak{Q}$, of all quasi-regular right ideals,[3] the radical, $\mathfrak{S}$, defined by Albert [4] (slightly generalized) and, for algebras with a unity element, the set $\mathfrak{R}$, of all h such that $g+h$ is regular for every regular g.[5]

The $\mathfrak{N}$-, $\mathfrak{Q}$-, and $\mathfrak{S}$-radicals are unaltered if a unity quantity is adjoined to $\mathfrak{A}$ (see Section **5**). Hence we consider alternative algebras with a unity and show, over a field containing more than two elements (this restriction enables us to use a theorem of Shoda which yields a very simple proof) that $\mathfrak{R}$ is an ideal such that the $\mathfrak{R}$-radical of $\mathfrak{A}-\mathfrak{R}$ is zero. The subsequent proof that $\mathfrak{N}=\mathfrak{R}$ provides a new and simpler proof of Zorn's results in this connection for the case of algebras.

Finally we show that $\mathfrak{N}$,$\mathfrak{Q}$, and $\mathfrak{S}$ are all identical, since after adjunction of a unity quantity they all coincide with $\mathfrak{R}$.[6]

1. The $\mathfrak{R}$-radical. By use of the Theorem of Artin, an alternative algebra may be characterized as an algebra $\mathfrak{A}$ obeying the following weakened form of the associative law: the sub-algebra $\mathfrak{A}<x,y>$ generated by an arbitrary pair of elements x and y of $\mathfrak{A}$ is associative.

An element x of $\mathfrak{A}$ is said to be *right quasi-regular* in case there is an element y in $\mathfrak{A}$ such that $x+y+xy=0$; y is then called a *right quasi-inverse* of x.

* Received September 5, 1947.

[1] Presented to the Society November 27, 1943, under the title "On the radical of a non-associative algebra."

[2] Zorn [1].

[3] Jacobson [1].

[4] Albert [1].

[5] Perlis [1].

[6] M. F. Smiley has shown in a paper presented to the American Mathematical Society on September 3, 1947 that Jacobson's definition of the radical also applies to alternative rings of possibly infinite dimensions, thus also generalizing Zorn's results [3] as well as our result on the Q-radical.

LEMMA 1. *If an element x of an alternative algebra $\mathfrak{A}$ has a right quasi-inverse y, y is also a left quasi-inverse, is unique, and is a polynomial in x. Hence x, y, and an arbitrary element z will always obey the associative law for multiplication.*

Since $\mathfrak{A} < x, y >$ is associative, the properties stated follow from familiar (and easily proved) properties of associative algebras. Now y is in $\mathfrak{A} < x, z >$, whence the final statement holds.

Right quasi-regularity is thus equivalent to quasi-regularity (in the case of alternative algebras), the same element y being at once the right and left quasi-inverse of x.

If $\mathfrak{A}$ has a unity element 1, x is quasi-regular if and only if $1 + x$ is regular. Then Lemma 1 is valid if "quasi-inverse" is replaced everywhere by "inverse." Note, also, that $\mathfrak{A} < x, y, 1 >$ is associative for every x and y of $\mathfrak{A}$.

LEMMA 2. *A product of regular elements g_1 and g_2 of an alternative algebra $\mathfrak{A}$ is regular. Conversely, if $g_1 g_2$ is regular, so are g_1 and g_2.*

For the inverses g_1^{-1} and g_2^{-1} of g_1 and g_2 lie in the associative sub-algebra $\mathfrak{A} < g_1, g_2, 1 > = \mathfrak{A} < g_1, g_2 >$, whence the product has inverse $g_2^{-1} g_1^{-1}$. Conversely, if $g_1 g_2$ has inverse g, g lies in $\mathfrak{A} < g_1, g_2 >$ so $(g_1 g_2) g = g_1 (g_2 g) = 1 = (g g_1) g_2$.

LEMMA 3. *If the alternative algebra $\mathfrak{A}$ has a unity element and the scalar field $\mathfrak{F}$ has at least three distinct elements, every element of $\mathfrak{A}$ is a sum of regular elements.*

Every x of $\mathfrak{A}$ is in an associative sub-algebra $\mathfrak{A} < x, 1 >$ in which Shoda's theorem [7] applies, yielding this lemma.

Hereafter it is assumed that $\mathfrak{F}$ is not the prime field of characteristic two unless $\mathfrak{F}$ is specified as being arbitrary.

Definition 1. If $\mathfrak{A}$ has a unity element, the $\mathfrak{R}$-*radical* of $\mathfrak{A}$ is the totality of elements r of $\mathfrak{A}$ such that $g + r$ is regular for every regular g of $\mathfrak{A}$.

THEOREM 1. *If $\mathfrak{A}$ is an alternative algebra with a unity element, the $\mathfrak{R}$-radical $\mathfrak{R}$ of $\mathfrak{A}$ is an ideal.*

It is clear that the set $\mathfrak{R}$ is a linear space. If g and g_i are any regular elements of $\mathfrak{A}$,

$$g + g_i r = g_i (g_i^{-1} g + r),$$

[7] Shoda [1], p. 254.

which is regular by a double application of Lemma 2. Hence $g_i r$ is in $\mathfrak{R}$. By Lemma 3 every a of $\mathfrak{A}$ has the form $a = \Sigma g_i$, $ar = \Sigma g_i r$. The fact that $\mathfrak{R}$ is a linear space then implies that ar is in $\mathfrak{R}$. Likewise $\mathfrak{R}$ contains every ra, $\mathfrak{R}$ is an ideal.

Under the hypotheses and notation of Theorem 1 the following result is valid.

THEOREM 2. *The $\mathfrak{R}$-radical of $\mathfrak{A} - \mathfrak{R}$ is zero.*

For let $[x]$ denote the coset modulo $\mathfrak{R}$ of x in $\mathfrak{A}$. If $[x]$ is any regular quantity of $\mathfrak{A} - \mathfrak{R}$, $[x]\,[y] = [xy] = [1]$ for some y of $\mathfrak{A}$, $xy = 1 + r$, r in $\mathfrak{R}$, xy is regular, and so are x and y by Lemma 2. Now let $[a]$ be in the $\mathfrak{R}$-radical of $\mathfrak{A} - \mathfrak{R}$ so $[x] + [a] = [x + a]$ is regular, and hence $x + a$ is regular. But x is an arbitrary regular quantity of $\mathfrak{A}$, so a is in $\mathfrak{R}$, $[a] = [0]$.

2. The $\mathfrak{N}$-radical. Our fundamental result is given by

THEOREM 3. *The set $\mathfrak{N}$ consisting of all properly nilpotent elements and the element* 0 *coincides with $\mathfrak{R}$ (and is called the $\mathfrak{N}$-radical of A).*

For if x is nilpotent of index t, $y = -x + x^2 + \cdots + (-1)^{t-1}x^{t-1}$ is the quasi-inverse of x. Let g be any regular element and u be in $\mathfrak{N}$. Then $x = g^{-1}u$ is nilpotent, hence quasi-regular by the opening remark, so $1 + g^{-1}u$ is regular. Then $g + u$ is regular by Lemma 2, so that u is in $\mathfrak{R}$, $\mathfrak{N} \leqq \mathfrak{R}$.

Conversely, every r of $\mathfrak{R}$ satisfies an equation

$$\alpha_0 + \alpha_1 r + \cdots + \alpha_t r^t = 0 \qquad (\alpha_i \text{ in } \mathfrak{F})$$

with $\alpha_t \neq 0$. If $\alpha_0 \neq 0$, r would have an inverse $-\alpha_0^{-1}(\alpha_1 + \cdots + \alpha_t r^{t-1})$. But if a regular element g is in $\mathfrak{R}$, so is $-g$, and $g + (-g)$ must be regular, which is impossible. Hence $\mathfrak{R}$ contains no regular elements and $\alpha_0 = 0$. Let j be the minimum subscript such that $\alpha_j \neq 0$; then

$$(\alpha_j + r_0)r^j = 0, \qquad r_0 = \alpha_{j+1}r + \cdots + \alpha_t r^{t-j},$$

r_0 is in $\mathfrak{R}$, $g = \alpha_j + r_0$ is regular, and $\mathfrak{A} < g, r >$ contains $g^{-1}(gr^j) = r^j = 0$. Thus r is nilpotent. Since $\mathfrak{R}$ is an ideal its elements are properly nilpotent, $\mathfrak{R} \leqq \mathfrak{N}$, $\mathfrak{R} = \mathfrak{N}$.

We have now shown that $\mathfrak{N}$ is an ideal, hence a nil ideal. Every nil right ideal $\mathfrak{B}$ of $\mathfrak{A}$ consists of 0 and certain properly nilpotent elements, so $\mathfrak{B} \leqq \mathfrak{N}$. Thus $\mathfrak{N}$ may be described as the sum of all nil right (left) ideals of $\mathfrak{A}$.

3. The $\mathfrak{Q}$-radical. We begin by making

Definition 2. The *$\mathfrak{Q}$-radical, Q, of* $\mathfrak{A}$ is the sum of all *quasi-regular right ideals* of $\mathfrak{A}$ (right ideals all of whose elements are quasi-regular).

LEMMA 4. *The $\mathfrak{Q}$-radical $\mathfrak{Q}$ of $\mathfrak{A}$ of Theorem* 1 *over an arbitrary field is a quasi-regular right ideal of* $\mathfrak{A}$.

Proof. It is clear that $\mathfrak{Q}$ is a right ideal. If $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ are any quasi-regular right ideals and $q_1 + q_2$ an arbitrary quantity of their sum $(\mathfrak{Q}_1, \mathfrak{Q}_2)$ q_i in $\mathfrak{Q}_i$, $g_1 = 1 + q_1$ is regular, $1 + q_1 + q_2 = g_1 + q_2 = (1 + q_2 g_1^{-1}) g_1$ in the associative algebra $\mathfrak{A} < q_1, q_2, 1 >$. Since $q_2 g_1^{-1}$ is in $\mathfrak{Q}_2$, $g_2 = 1 + q_2 g_1^{-1}$ is regular and so is $g_2 g_1 = 1 + q_1 + q_2$ by Lemma 2. Hence $q_1 + q_2$ is quasi-regular. An obvious induction proves that the sum of any finite number of quasi-regular right ideals is quasi-regular, and thus every element of $\mathfrak{Q}$ is quasi-regular.

THEOREM 4. *The algebra $\mathfrak{A}$ of Theorem* 1 *has $\mathfrak{Q}$-radical $\mathfrak{Q} = \mathfrak{R} = \mathfrak{N}$.*

Proof. For each regular element g of $\mathfrak{A}$ and each q of $\mathfrak{Q}$, qg^{-1} is in $\mathfrak{Q}$. Hence $1 + qg^{-1}$ is regular and so is $g + q$. Thus $\mathfrak{Q} \leqq \mathfrak{R}$. But $1 + r$ is regular for each r of $\mathfrak{R}$, r is quasi-regular; $\mathfrak{R}$ is a quasi-regular ideal; $\mathfrak{R} \leqq \mathfrak{Q}$; $\mathfrak{R} = \mathfrak{Q}$.

4. The $\mathfrak{S}$-radical. We make

Definition 3. *The $\mathfrak{S}$-radical $\mathfrak{S}$ of* $\mathfrak{A}$ is the intersection of all ideals[8] $\mathfrak{B}$ of $\mathfrak{A}$ such that $\mathfrak{A} - \mathfrak{B}$ is a direct sum of simple algebras.

THEOREM 5. *The $\mathfrak{S}$-radical of the algebra $\mathfrak{A}$ of Theorem* 1 *is* $\mathfrak{S} = \mathfrak{R} = \mathfrak{Q} = \mathfrak{N}$.

Proof. Zorn has proved[9] that $\mathfrak{A} - \mathfrak{N}$ is a direct sum of simple algebras. Therefore, $\mathfrak{S} \leqq \mathfrak{N}$. Since 1 is not in $\mathfrak{N}$, $\mathfrak{A} > \mathfrak{N} \geqq \mathfrak{S}$.

Since

$$\mathfrak{A} - \mathfrak{S} = \mathfrak{B}_1 \oplus \cdots \oplus \mathfrak{B}_n$$

---

[8] Such ideals always exist when $\mathfrak{A}$ has a unity element for all non-associative algebras. See Albert [1], p. 897.

[9] Zorn [2], p. 1.

where the $\mathfrak{B}_i$ are[10] simple algebras with unity elements e_i, each element x of $\mathfrak{N}$ corresponds to a coset

$$[x] = [x_1] + \cdots + [x_n], \qquad [x_i] \text{ in } \mathfrak{B}_i$$

where the $[x_i]$ are properly nilpotent in $\mathfrak{B}_i$ and $\mathfrak{A} - \mathfrak{S}$. As $[x]$ varies over $\mathfrak{N}$, $[x_i]$ varies over an ideal $\mathfrak{N}_i$ of $\mathfrak{B}_i$, so $\mathfrak{N}_i = 0$ or $\mathfrak{N}_i = \mathfrak{B}_i$. In the latter case e_i is in $\mathfrak{N}_i$ whereas e_i is not nilpotent. Hence each $\mathfrak{N}_i = 0$, $\mathfrak{N} - \mathfrak{S} = 0$, $\mathfrak{N} \leqq \mathfrak{S} \leqq \mathfrak{N}$, $\mathfrak{N} = \mathfrak{S}$.

That $\mathfrak{S} = \mathfrak{N}$ may also be proved by use of the fact that $\mathfrak{N} = \mathfrak{Q}$. We already have $\mathfrak{S} \leqq \mathfrak{N} = \mathfrak{Q} < \mathfrak{A}$. Now $\mathfrak{Q} - \mathfrak{S}$ is a quasi-regular ideal of $\mathfrak{A} - \mathfrak{S} = \mathfrak{B}_1 \oplus \cdots \oplus \mathfrak{B}_n$. For each x in $\mathfrak{Q}$, $[x] = [x_1] + \cdots + [x_n]$, where $[x_i]$ in $\mathfrak{B}_i$ varies over a quasi-regular ideal $\mathfrak{Q}_i$ of $\mathfrak{B}_i$ as x varies over $\mathfrak{Q}$. If $\mathfrak{Q}_i$ is not 0, it must contain $-e_i$ (where e_i is the unity of $\mathfrak{B}_i$). But $-e_i$ is obviously not quasi-regular, so each $\mathfrak{Q}_i = 0$, $\mathfrak{Q} - \mathfrak{S} = 0$, $\mathfrak{Q} \leqq \mathfrak{S} \leqq \mathfrak{Q}$.

5. Alternative algebras without unity elements. The preceding work has been concerned with an alternative algebra $\mathfrak{A}$ with a unity element. The definition of the $\mathfrak{R}$-radical requires a unity element, but the definitions of the $\mathfrak{N}$-, $\mathfrak{Q}$-, and $\mathfrak{S}$-radicals[11] are meaningful for an arbitrary alternative algebra $\mathfrak{A}_0$. These radicals of $\mathfrak{A}_0$ shall be denoted by $\mathfrak{N}_0$, $\mathfrak{Q}_0$, and $\mathfrak{S}_0$, and we shall show that they coincide with $\mathfrak{N}$, $\mathfrak{Q}$, and $\mathfrak{S}$, respectively, for the algebra $\mathfrak{A} = (\mathfrak{A}_0, 1)$ obtained by adjoining 1 to $\mathfrak{A}_0$.

THEOREM 6. *The $\mathfrak{N}$-, $\mathfrak{Q}$-, and $\mathfrak{S}$-radicals of any alternative algebra $\mathfrak{A}_0$ over an arbitrary field $\mathfrak{F}$ coincide, respectively, with the corresponding radicals of the algebra $\mathfrak{A} = (\mathfrak{A}_0, 1)$.*

For if $a = \alpha + a_0$ is in $\mathfrak{N}$, α in $\mathfrak{F}$, a_0 in $\mathfrak{A}_0$, $a^t = 0 = \alpha^t + a_1$, a_1 in $\mathfrak{A}_0$, $\alpha^t = 0 = \alpha$, $\mathfrak{N} \leqq \mathfrak{A}_0$, $\mathfrak{N}$ consists of properly nilpotent elements of $\mathfrak{A}_0$, $\mathfrak{N} \leqq \mathfrak{N}_0$.

The converse is more difficult. If x is any non-zero quantity of $\mathfrak{N}_0$, x generates an associative sub-algebra $\mathfrak{B}$ of $\mathfrak{A}_0$, x is properly nilpotent in $\mathfrak{B}$, whence, by familiar results for associative algebras, x is nilpotent.

If, in addition, y is any nilpotent quantity of $\mathfrak{A}_0$, $\mathfrak{B}_0 = \mathfrak{A}_0 < x, y >$ is associative, x is in the radical $\mathfrak{R}_0$ of $\mathfrak{B}_0$, $(x + y)^t = x^t + x_2 + \cdots + x_s + y^t$ with x_i in $\mathfrak{R}_0$ and $y^t = 0$ for sufficiently large t. Thus $(x + y)^t = z$ in $\mathfrak{R}_0$, $z^u = 0$, $(x + y)^{tu} = 0$. With this result we are able to prove that x is properly nilpotent in $\mathfrak{A}$.

---

[10] Albert [2], p. 709.

[11] If the $\mathfrak{S}$-radical of $\mathfrak{A}_0$ as originally defined does not exist, we define it to be $\mathfrak{A}_0$.

Let $\alpha a = \alpha(1 + a_0)$ be any quantity of $\mathfrak{A}$ not in $\mathfrak{A}_0$, a_0 in $\mathfrak{A}_0$, $\alpha \neq 0$ in $\mathfrak{F}$. Then for x in $\mathfrak{R}_0$, $xa = x + xa_0 = x + y$ is in $\mathfrak{A}_0$, y is nilpotent, and so is $x + y$ by the result above. But then $x \cdot \alpha a = \alpha(x + y)$ is nilpotent, x is properly nilpotent in $\mathfrak{A}$, $\mathfrak{R}_0 \leqq \mathfrak{R}$, $\mathfrak{R}_0 = \mathfrak{R}$.

It is easy to prove now the following result (if the field $\mathfrak{F}$ contains at least three elements):

COROLLARY. *The $\mathfrak{R}$-radical of an alternative algebra $\mathfrak{A}_0$ is the sum of all nil right (left) ideals of $\mathfrak{A}_0$.*

To continue with the proof of Theorem 6 by showing $\mathfrak{Q}_0 = \mathfrak{Q}$, note first that the proof of Lemma 4 may be used to show that $\mathfrak{Q}_0$, which is a right ideal of $\mathfrak{A}_0$ and of $\mathfrak{A}$, is quasi-regular in $\mathfrak{A}$, so $\mathfrak{Q}_0 \leqq \mathfrak{Q}$. Let $a = \alpha + a_0$ be in $\mathfrak{Q}$, α in $\mathfrak{F}$, a_0 in $\mathfrak{A}_0$. If $\alpha \neq 0$, $-\alpha^{-1}a = -1 + b_0$ is in the ideal $\mathfrak{Q}$, b_0 in $\mathfrak{A}_0$. But then $1 + (-1 + b_0) = b_0$ is regular in $\mathfrak{A}$, which is impossible. Hence $\alpha = 0$, $\mathfrak{Q} \leqq \mathfrak{A}_0$, $\mathfrak{Q} \leqq \mathfrak{Q}_0$.

To show that $\mathfrak{S} = \mathfrak{S}_0$ we prove the following more general result.

THEOREM 7. *Let $\mathfrak{A}_0$ be an arbitrary non-associative algebra and $\mathfrak{A} = (\mathfrak{A}_0, 1)$. If $\mathfrak{S}$ and $\mathfrak{S}_0$ are the $\mathfrak{S}$-radicals of $\mathfrak{A}$ and $\mathfrak{A}_0$, respectively, $\mathfrak{A}_0 \geqq \mathfrak{S}$. If $\mathfrak{A}_0 > \mathfrak{S}$, then*

(a) $\mathfrak{S} \geqq \mathfrak{S}_0$.

(b) *$\mathfrak{A}_0 - \mathfrak{S}$ has a unity element.*

(c) *$\mathfrak{S} = \mathfrak{S}_0$ if and only if $\mathfrak{A}_0 - \mathfrak{S}_0$ has a unity element.*

Since $\mathfrak{A} - \mathfrak{A}_0$ is simple, it follows that $\mathfrak{A}_0 \geqq \mathfrak{S}$. Suppose $\mathfrak{A}_0 > \mathfrak{S}$. The algebra $\mathfrak{A} - \mathfrak{S}$ has the forms

$$\mathfrak{A} - \mathfrak{S} = (\mathfrak{A}_0 - \mathfrak{S}, 1) = \mathfrak{A}_1 \oplus \cdots \oplus \mathfrak{A}_n$$

where the $\mathfrak{A}_i$ are simple sub-algebras of $\mathfrak{A} - \mathfrak{S}$ and $\mathfrak{A}_0 - \mathfrak{S}$ is an ideal of $\mathfrak{A} - \mathfrak{S}$. Then [12] $\mathfrak{A}_0 - \mathfrak{S} = \mathfrak{B}_1 \oplus \cdots \oplus \mathfrak{B}_n$ where each $\mathfrak{B}_i$ is an ideal of $\mathfrak{A}_i$, so $\mathfrak{B}_i = 0$ or $\mathfrak{B}_i = \mathfrak{A}_i$ and at least one $\mathfrak{B}_i = \mathfrak{A}_i$ since $\mathfrak{A}_0 > \mathfrak{S}$. It follows that $\mathfrak{S} \geqq \mathfrak{S}_0$. Since each $\mathfrak{A}_i$ has [13] a unity element, (b) is true.

If $\mathfrak{A}_0 - \mathfrak{S}_0$ has a unity element $[e]$,

$$\begin{aligned}\mathfrak{A} - \mathfrak{S}_0 &= ([1], \mathfrak{A}_0 - \mathfrak{S}_0) = ([1 - e], \mathfrak{A}_0 - \mathfrak{S}_0) \\ &= ([1 - e]) \oplus (\mathfrak{A}_0 - \mathfrak{S}_0)\end{aligned}$$

since $[1 - e]$ is an idempotent orthogonal to the unity element of $\mathfrak{A}_0 - \mathfrak{S}_0$. But $([1 - e])$ is a field and $\mathfrak{A}_0 - \mathfrak{S}_0$ is a direct sum of simple sub-algebras.

---

[12] Albert [2], p. 710.
[13] Albert [2], p. 709.

Thus $\mathfrak{S}_0$ is one of the ideals of $\mathfrak{A}$ whose intersection is $\mathfrak{S}$, $\mathfrak{S}_0 \geqq \mathfrak{S}$, $\mathfrak{S}_0 = \mathfrak{S}$. The converse is trivial in view of (b), and the proof of (c) is complete.

Apply Theorem 7 to alternative algebras, note first that every simple alternative algebra has[14] a unity element. Consider the case $\mathfrak{A}_0 > \mathfrak{S}$. Then $\mathfrak{A}_0 - \mathfrak{S}_0$ has a unity element, whence the theorem implies that $\mathfrak{S}_0 = \mathfrak{S}$. Now let $\mathfrak{A}_0 = \mathfrak{S}$ and suppose $\mathfrak{A}_0 > \mathfrak{S}_0$. Then the last paragraph above shows that $\mathfrak{S}_0 \geqq \mathfrak{S} = \mathfrak{A}_0$, $\mathfrak{S}_0 = \mathfrak{A}_0$, a contradiction. Hence $\mathfrak{A}_0 = \mathfrak{S}_0 = \mathfrak{S}$.

SYRACUSE UNIVERSITY
AND
PURDUE UNIVERSITY.

BIBLIOGRAPHY.

A. A. Albert:

[1] "The radical of a non-associative algebra," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 891-897.

[2] "Non-associative algebras II. New simple algebras," *Annals of Mathematics*, vol. 43 (1942), pp. 708-723.

N. Jacobson:

[1] "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

Sam Perlis:

[1] "A characterization of the radical of an algebra," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 128-132.

K. Shoda:

[1] "Über die Galoissche Theorie der halbeinfachen hypercomplexen Systeme," *Mathematische Annalen*, vol. 107 (1932), pp. 252-258.

Max Zorn:

[1] "Alternative rings and related questions I: Existence of the radical," *Annals of Mathematics*, vol. 43 (1941), pp. 676-685.

[2] "Theorie der alternativen Ringe," *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, vol. 8 (1930), pp. 123-147.

[14] Zorn [2], p. 147.

SOME REMARKS ON THE DISTRIBUTION OF SEQUENCES OF REAL NUMBERS.*

By F. A. BEHREND.

Let

$$x_1, x_2, x_3, \cdots \tag{1}$$

be a sequence of real numbers and $\phi_n(x) = (1/n) \sum_{\substack{\nu=1 \\ x_\nu < x}}^{n} 1$. (1) is said to have an asymptotic distribution $\phi(x)$ if $\phi(x)$ is a distribution function[1] and if $\phi_n(x) \to \phi(x)$ for all continuity points of $\phi(x)$.

This paper is concerned with the investigation of a sufficient condition for (1) to permit the term by term integration

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)\, d\phi_n(x) = \int_{-\infty}^{\infty} f(x)\, d\phi(x) \tag{2}$$

where $f(x)$ is any continuous monotonic function, and with the modification of given sequences in such a way as to make them satisfy the condition obtained.

By the Helly-Bray theorem,[2] (2) is valid if (1) is a bounded sequence. It will be shown that (2) remains valid if (1) is unbounded but does not increase too rapidly in a sense to be specified below (Definition (4), Theorem I). Such a sequence will be called *normal.* It will also be shown that any sequence with an asymptotic distribution can be normalized by a simple modification which does not alter its distribution function. In the case of a continuous distribution function $\phi(x)$ the modification consists in the suppression of a subsequence, the number of terms suppressed from $x_1, \cdots, x_n$ being $o(n)$ (Theorem II). This is a best possible result as, conversely, the

---

* Received July 19, 1947.

[1] A monotonic non-decreasing function is called a distribution function if $\lim_{x\to-\infty} \phi(x) = 0$, $\lim_{x\to\infty} \phi(x) = 1$.

[2] H. E. Bray, "Elementary properties of the Stieltjes integral," *Annals of Mathematics*, vol. 20 (1919), pp. 177-186; E. Helly, "Über lineare Funktionaloperationen," *Sitzungsberichte der Naturwissenschaftlichen Klasse der Kaiserlichen Akademie der Wissenschaften Wien*, vol. 121 (1912), part IIa, numbers I to X, pp. 265-297.

validity of (2) can be destroyed by the insertion of such a sequence.[3] In the case of a discontinuous $\phi(x)$ an intermediate step is necessary: x_n is replaced by a sequence $x_n + \alpha_n$ where $\alpha_n \to 0$, and then a subsequence of $x_n + \alpha_n$ is suppressed (Theorems III, II′).

1. Sequences with continuous distribution. *Normal sequences.* Let $\mu_n = \min(x_1, \cdots, x_n)$, $M_n = \max(x_1, \cdots, x_n)$ and

$$\epsilon_n = \overline{\mathrm{fin}}\, |\phi(x) - \phi_n(x)|. \tag{3}$$

The sequence x_n will be called *normal* if

$$\epsilon_n = O(\min(\phi(\mu_n), 1 - \phi(M_n))). \tag{4}$$

As this condition may be interpreted as giving a lower bound for μ_n and an upper bound for M_n, it restricts the rapidity of increase of the sequence x_n in terms of the approximation ϵ_n. It is the sharpest such condition that can be imposed; for, considering $\lim(\phi(x) - \phi_n(x))$ as $x \to \mu_n - 0$ or $M_n + 0$, it is seen that $\epsilon_n \geq \max(\phi(\mu_n), 1 - \phi(M_n))$.

Condition for term by term integration.

THEOREM I. *If x_n is normal, $f(x)$ continuous and monotonic, then*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f d\phi_n = \int_{-\infty}^{\infty} f d\phi \tag{2}$$

provided that the integral on the right hand side converges.

Proof. As (2) is valid for f if it holds for $-f$ or $f +$ const. it may be assumed that f is non-decreasing and that $f(x_1) = 0$. Then $f(x) \geq 0$ for $x \geq x_1$, and $f(x) \leq 0$ for $x \leq x_1$. Consider first the range (x_1, ∞).

(i) If M_n is bounded, $M_n < M$, say, then $\phi(x) = \phi_n(x) = 1$ for $x \geq M$ and

$$\int_{x_1}^{\infty} f d\phi - \int_{x_1}^{\infty} f d\phi_n = \int_{x_1}^{M} f d(\phi - \phi_n) \to 0,$$

by the Helly-Bray theorem.

---

[3] E. g., if the term $2cn$ is inserted after x_{n^2}, the $\lim\limits_{n\to\infty} \int_{-\infty}^{\infty} x d\phi_n$ is increased by c, whereas $\int_{-\infty}^{\infty} x d\phi$ remains unaltered.

(ii) If M_n is unbounded, $M_n \to \infty$, then $\phi_n(x) = 1$ for $x > M_n$ and

$$\int_{x_1}^{\infty} f d\phi - \int_{x_1}^{\infty} f d\phi_n = \int_{M_n}^{\infty} f d\phi + \int_{x_1}^{M_n+0} f d(\phi - \phi_n)$$
$$= \int_{M_n}^{\infty} f d\phi - f(M_n)(1 - \phi(M_n)) - \int_{x_1}^{M_n+0} (\phi - \phi_n) df \to 0;$$

for, $\int_{M_n}^{\infty} f d\phi \to 0$, by the assumed convergence of $\int_{x_1}^{\infty} f d\phi$; and the other terms tend to 0 as

$$0 \leq f(M_n)(1 - \phi(M_n) \leq \int_{M_n}^{\infty} f d\phi$$

and

$$\left| \int_{x_1}^{M_n+0} (\phi - \phi_n) df \right| \leq \epsilon_n f(M_n), \text{ by (3)}$$
$$= O(f(M_n)(1 - \phi(M_n))), \text{ by (4)}.$$

As the range $(-\infty, x_1)$ may be treated similarly, I is established.[4]

Note. I remains valid if one of $\int_{-\infty}^{x_1} f d\phi$, $\int_{x_1}^{\infty} f d\phi$ diverges. This is an immediate consequence of Fatou's lemma.

Normalization of a sequence.

THEOREM II. *By the suppression of a suitably chosen subsequence any sequence x_n with continuous distribution $\phi(x)$ can be transformed into a normal sequence y_m with the same distribution.*

Proof. As $\phi(x)$ increases continuously from 0 to 1 real numbers s_k, S_k exist such that

(5) $$\phi(s_k) = 1 - \phi(S_k) = 2^{-k}, \qquad k = 1, 2, 3, \cdots$$

where

(6) $$\cdots < s_3 < s_2 < s_1 = S_1 < S_2 < S_3 < \cdots.$$

Define a sequence of integers n_k $(k = 0, 1, 2, \cdots)$ such that $n_0 = 0$ and, for $k > 0$,

(7) $$n_k > 2^k n_{k-1}$$

and

(8) $$\epsilon_n < 2^{-k} \text{ for } n \geq n_k,$$

[4] The proof is divided into the two parts (i), (ii) with a view to the extension below to discontinuous distributions ϕ. The division is unnecessary for continuous ϕ, as in this case $\int_{M_n}^{\infty} f\, d\phi \to 0$ whether or not M_n is bounded.

where ϵ_n is defined by (3). (8) can be satisfied as $\epsilon_n \to 0$ by the following theorem due to Pólya [5]: If ϕ_n, ϕ are distribution functions, $\phi_n \to \phi$, and if ϕ is continuous, then the convergence is uniform.

The sequence y_m is now obtained as follows: for each k $(k = 1, 2, 3, \cdots)$ suppress those elements x_ν of the original sequence for which

$$n_{k-1} < \nu \leq n_k \text{ and } x_\nu < s_k \text{ or } x_\nu \geq S_k. \tag{9}$$

Let $y_1, \cdots, y_m$ be those of $x_1, \cdots, x_n$ which have not been suppressed. $m = m(n)$ may be estimated as follows: let k be such that $n_k < n \leq n_{k+1}$; then

$$\begin{aligned} m(n) &= m(n_{k-1}) + n_k\{\phi_{n_k}(S_k) - \phi_{n_k}(s_k)\} - n_{k-1}\{\phi_{n_{k-1}}(S_k) - \phi_{n_{k-1}}(s_k)\} \\ &\qquad + n\{\phi_n(S_{k+1}) - \phi_n(s_{k+1})\} - n_k\{\phi_{n_k}(S_{k+1}) - \phi_{n_k}(s_{k+1})\} \\ &= O(n_{k-1}) + n_k\{\phi_{n_k}(S_k) - \phi_{n_k}(s_k) - \phi_{n_k}(S_{k+1}) + \phi_{n_k}(s_{k+1})\} \\ &\qquad + n\{\phi_n(S_{k+1}) - \phi_n(s_{k+1})\}. \end{aligned} \tag{10}$$

By (7), the first term is $O(n_{k-1}) = O(2^{-k}n_k) = O(2^{-k}n)$. If, in the other terms, ϕ_{n_k}, ϕ_n are replaced by ϕ the errors made are at most ϵ_{n_k}, ϵ_n which are $O(2^{-k})$ by (8). Substituting the values of ϕ given by (5) we get

$$m(n) = n\{1 + O(2^{-k})\}, \text{ i.e., } (n - m)/n = O(2^{-k}). \tag{11}$$

Let $\psi_m(x)$ be the distribution of $y_1, \cdots, y_m$; then

$$\begin{aligned} |\phi(x) - \psi_m(x)| &\leq |\phi(x) - \phi_n(x)| + |\phi_n(x) - (m/n)\psi_m(x)| + |(m/n)\psi_m(x) - \psi_m \\ &\leq \epsilon_n + (1/n)|n\phi_n(x) - m\psi_m(x)| + ((n-m)/n)\psi_m(x) \\ &\leq \epsilon_n + (n-m)/n + (n-m)/n = O(2^{-k}) \end{aligned}$$

by (8) and (11). As $m \to \infty$, $n \to \infty$ and $k \to \infty$; hence $\psi_m(x) \to \phi(x)$ for all x, i.e., the sequence y_m has the distribution $\phi(x)$. By (9) and (6), all elements $< s_{k+1}$ or $\geq S_{k+1}$ of the original sequence are suppressed; thus $s_{k+1} \leq y_\mu < S_{k+1}$ for $\mu = 1, 2, \cdots, m$, whence

$$s_{k+1} \leq \min(y_1, \cdots, y_m) = \nu_m, \qquad S_{k+1} > \max(y_1, \cdots, y_m) = \mathrm{N}_m,$$

and

$$2^{-k} = 2\phi(s_{k+1}) = 2(1 - \phi(S_{k+1})) = O(\min(\phi(\nu_m), 1 - \phi(\mathrm{N}_m))),$$

and by (12)

[5] G. Pólya, "Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem," *Mathematische Zeitschrift*, vol. 8 (1920), pp. 171-181, Satz I.

$$\delta_m = \overline{\text{fin}}\, |\, \phi(x) - \psi_m(x)\, | = O(2^{-k}) = O(\min\, (\phi(\nu_m), 1 - \phi(\mathrm{N}_m))),$$

i. e. the sequence y_m is normal.

2. Sequences with discontinuous distribution. In the definition of a normal sequence, formulae (3) and (4) have to be replaced by

$$(3') \qquad \epsilon_n = \overline{\text{fin}}\, |\, \phi(x-0) - \phi_n(x-0)\, | = \overline{\text{fin}}\, |\, \phi(x+0) - \phi_n(x+0)\, |\,,$$

$$(4') \qquad \epsilon_n = O(\min\, (\phi(\mu_n + 0), 1 - \phi(\mathrm{M}_n - 0))).$$

Theorem I can be obtained by obvious modifications of the above proof, but the proof of Theorem II is no longer valid as Pólya's theorem is not available in the case of a discontinuous distribution and ϵ_n is not, in general, a null-sequence. This case may however be treated using a generalization of Pólya's result:

LEMMA 1. *If $\phi_n(x)$, $\phi(x)$ are distribution functions and if, for all x,*

$$(13) \qquad \lim_{n\to\infty} \phi_n(x-0) = \phi(x-0) \text{ and } \lim_{n\to\infty} \phi_n(x+0) = \phi(x+0),$$

then (13) *holds uniformly for all x, i. e. $\epsilon_n \to 0$.*

Lemma 1 may easily be proved on the lines of Pólya's original proof.[6] A sequence satisfying (13) will be called *regular.*

THEOREM III. *If x_n is any sequence with asymptotic distribution $\phi(x)$ then a null sequence α_n exists such that $z_n = x_n + \alpha_n$ is a regular sequence with the same distribution.*

The proof of Theorem III rests on the

LEMMA 2. *If ξ is a discontinuity of the distribution function $\phi(x)$ of x_n, and $\eta = \phi(\xi + 0) - \phi(\xi - 0)$, then a subsequence y_m of x_n exists such that*

$$(14) \qquad \lim_{m\to\infty} y_m = \xi$$

and

$$(15) \qquad \lim_{n\to\infty} m(n)/n = \eta,$$

where $m(n)$ is the number of y's selected from $x_1, \cdots, x_n$.

Proof of Lemma 2. The procedure is similar to that used in the proof

---

[6] See F. A. Behrend, "The uniform convergence of sequences of monotonic functions," *Proceedings of the Royal Society of New South Wales*, vol. 81 (1948), pp. 167-168.

of II. For $k = 1, 2, 3, \cdots$ continuity points s_k, S_k of $\phi(x)$ are chosen such that.

$$(5') \qquad 0 \leq \phi(S_k) - \phi(s_k) - \eta < 2^{-k}$$

where

$$(6') \qquad s_k < \xi < S_k \text{ and } s_k \to \xi - 0,\ S_k \to \xi + 0.$$

Integers $n_0 = 0, n_1, n_2, \cdots$ are defined such that, for $k = 1, 2, 3, \cdots$,

$$(7) \qquad n_k > 2^k n_{k-1}$$

and

$$(8') \qquad |\phi(s_k) - \phi_n(s_k)| < 2^{-k}, \quad |\phi(S_k) - \phi_n(S_k)| < 2^{-k} \text{ for } n \geq n_k.$$

The sequence y_m is determined as follows: for $k = 1, 2, 3, \cdots$ all those x_ν are selected for which

$$(9') \qquad n_{k-1} < \nu \leq n_k \text{ and } s_k \leq x_\nu < S_k.$$

(14) is an immediate consequence of (6′) and (9′). $m(n)$ is given by (10), and using (7) and (8′), (5′) instead of (8), (5), we get

$$(11') \qquad m(n) = n(\eta + O(2^{-k})),$$

whence (15).

Proof of Theorem III. Put $\eta(x) = \phi(x+0) - \phi(x-0)$, and let the discontinuities of $\phi(x)$ be, in any order, $\xi_1, \xi_2, \xi_3, \cdots$. By Lemma 2, there exists to any ξ_k $(k = 1, 2, 3, \cdots)$ a subsequence

$$(16) \qquad x_{p_{k1}}, x_{p_{k2}}, x_{p_{k3}}, \cdots$$

such that

$$(17) \qquad \lim_{m\to\infty} x_{p_{km}} = \xi_k$$

and

$$(18) \qquad \lim_{n\to\infty} m(k,n)/n = \eta(\xi_k),$$

where $m(k,n)$ is the number of elements of (16) selected from $x_1, \cdots, x_n$. Omitting, if necessary, a finite number of terms, (16) may be so chosen that

$$(19) \qquad |\xi_k - x_{p_{km}}| < 2^{-k} \qquad \text{for } m = 1, 2, 3, \cdots$$

and

$$(20) \qquad p_{km} \neq p_{k'm'} \qquad \text{for } k' < k \text{ and all } m, m',$$

so that the sequences (16) $(k = 1, 2, 3, \cdots)$ are disjunct subsequences of x_n. z_n is now defined by

$$(21) \qquad z_n = x_n \text{ if } n \neq p_{km} \text{ for all } k \text{ and } m,$$
$$z_n = \xi_k \text{ if } n = p_{km}.$$

It will first be shown that $\alpha_n = z_n - x_n \to 0$. To any $\epsilon > 0$, k exists such that $2^{-k} < \epsilon$.

(i) If $n \neq p_{lm}$ for all l, m, then $\alpha_n = z_n - x_n = x_n - x_n = 0$.

(ii) If $n = p_{lm}$ and $l \geqq k$, then $|\alpha_n| = z_n - x_n| = |\xi_l - x_{p_{lm}}| < 2^{-k} < \epsilon$, by (19).

(iii) If $n = p_{lm}$ and $l < k$, then $|\alpha_n| = |\xi_l - x_{p_{lm}}| < \epsilon$, provided that

$$n = p_{lm} > N(\epsilon).$$

Hence $\alpha_n \to 0$.

Let $\psi_n(x)$ be the distribution of $z_1, \cdots, z_n$. For each x

$$\psi_n(x+0) - \psi_n(x-0) \geqq \eta(x) + o(1). \tag{22}$$

For, if x is a continuity point of $\phi(x)$, $\eta(x) = 0$ and (22) is trivial; if $x = \xi_k$, then $z_1, \cdots, z_n$ contain at least $m(k,n)$ terms ξ_k, whence

$$\psi_n(x+0) - \psi_n(x-0) \geqq m(k,n)/n = \eta(\xi_k) + o(1),$$

by (18).

Let $\epsilon > 0$, and $N(\epsilon)$ be such that $|\alpha_n| < \epsilon$ for $n > N(\epsilon)$. At most $N(\epsilon)$ of the z_ν differ from the corresponding x_ν by more than ϵ; hence,

$$\phi_n(x-2\epsilon) - N(\epsilon)/n \leqq \psi_n(x-\epsilon) \leqq \psi_n(x-0)$$

and

$$\psi_n(x+0) \leqq \psi_n(x+\epsilon) \leqq \phi_n(x+2\epsilon) + N(\epsilon)/n,$$

whence

$$\begin{aligned}\phi_n(x-2\epsilon) - N(\epsilon)/n \leqq \psi_n(x-0) &= \psi_n(x+0) - \{\psi_n(x+0) - \psi_n(x-0)\} \\ &\leqq \phi_n(x+2\epsilon) - \eta(x) + o(1) + N(\epsilon)/n,\end{aligned}$$

by (22). Letting $n \to \infty$,

$$\phi(x-2\epsilon) \leqq \underline{\lim}\, \psi_n(x-0) \leqq \overline{\lim}\, \psi_n(x-0) \leqq \phi(x+2\epsilon) - \eta(x),$$

provided that $x - 2\epsilon$, $x + 2\epsilon$ are continuity points of $\phi(x)$; and as this condition may be satisfied for arbitrarily small ϵ,

$$\phi(x-0) \leqq \underline{\lim}\, \psi_n(x-0) \leqq \overline{\lim}\, \psi_n(x-0) \leqq \phi(x+0) - \eta(x) = \phi(x-0),$$

i. e.,

$$\lim_{n\to\infty} \psi_n(x-0) = \phi(x-0), \text{ and similarly, } \lim_{n\to\infty} \psi_n(x+0) = \phi(x+0). \tag{23}$$

It follows from (23) that $\psi_n(x) \to \phi(x)$ for all continuity points of $\phi(x)$; hence, the sequence $z_n = x_n + \alpha_n$ has the distribution $\phi(x)$ and is regular, by (23).

THEOREM II′. *By the suppression of a suitably chosen subsequence any regular sequence x_n with distribution $\phi(x)$ can be transformed into a normal sequence y_m with the same distribution.*

Proof. The procedure is modelled on that of the proof of Theorem II with the following modifications. s_k, S_k are chosen such that

$$(5'') \quad \phi(s_k - 0) \leq 2^{-k} \leq \phi(s_k + 0), \; 1 - \phi(S_k + 0) \leq 2^{-k} \leq 1 - \phi(S_k - 0),$$

where

$$(6'') \qquad \cdots \leq s_3 \leq s_2 \leq s_1 = S_1 \leq S_2 \leq S_3 \leq \cdots .$$

The sequence n_k is defined to satisfy (7) and (8), and (9) is replaced by

$$(9'') \qquad n_{k-1} < \nu \leq n_k \text{ and } x_\nu < s_k \text{ or } x_\nu > S_k.$$

In (10) the S have to be replaced by $S + 0$ and the s by $s - 0$, and (11) follows using (7), (8) and the estimates $\phi(s_k - 0) = O(2^{-k})$, $\phi(S_k + 0) = 1 + O(2^{-k})$ which hold by (5″). (12) holds with x replaced by $x - 0$ or $x + 0$, which shows that the sequence y_m is regular; and using that, by (5″),

$$\begin{aligned} 2^{-k} &\leq 2 \min\left(\phi(s_{k+1} + 0), 1 - \phi(S_{k+1} - 0)\right) \\ &\leq 2 \min\left(\phi(\nu_m + 0), 1 - \phi(\mathrm{N}_m - 0)\right), \end{aligned}$$

the sequence is shown to be normal.

UNIVERSITY OF MELBOURNE.

NONLINEAR HYPERBOLIC DIFFERENTIAL EQUATIONS FOR FUNCTIONS OF TWO INDEPENDENT VARIABLES.*

By K. O. Friedrichs.

The present paper is concerned with uniqueness, existence, and differentiability of solutions of the initial value problem of systems of hyperbolic differential equations for functions of two independent variables. Specifically we shall prove that a system of quasi-linear hyperbolic differential equations

$$\sum_{n=1}^{N} \left(a^{mn}\frac{\partial u^n}{\partial x} + b^{mn}\frac{\partial u^n}{\partial y}\right) = g^m, \qquad m = 1, \cdots, N,$$

possesses in a neighborhood $\mathfrak{R}$ of a section $\mathfrak{I}$ of the x-axis a unique solution $\{u^1(x,y), \cdots, u^N(x,y)\}$ which assumes given initial values $\{\bar{u}^1(x), \cdots, \bar{u}^N(x)\}$ on $\mathfrak{I}$, provided these initial functions possess continuous second derivatives with respect to x, and the coefficients a^{mn}, b^{mn}, and g^m possess continuous second derivatives with respect to $x, y, u^1, \cdots, u^N$. The solution is then proved to possess continuous second derivatives with respect to x and y. If the $\bar{u}^n$, a^{mn}, b^{mn}, g^m possess continuous third derivatives, the same is true for the solution $\{u^n\}$.

Before establishing these statements we shall prove theorems for linear equations and for "semi-linear" equations, i. e., equations whose coefficients a^{mn}, b^{mn} do not depend on $u^1, \cdots, u^N$, while the g^m may depend nonlinearly on these variables. In these theorems we shall impose weaker differentiability conditions than for quasi-linear equations.

We emphasize the fact that always the same properties will be established for the solutions that are required of the initial data, so that the values of the solution on any line $y = y_1 > 0$, $y_1 = \text{const.}$, could be used as initial data for a continuation of the solution. Properties of initial data of this character may be called "persistent" or "continuable."[1] Thus we can say, for example, that the property of initial data for the solutions of quasi-linear hyperbolic systems of first order to have continuous second (or third) derivatives is persistent.

The investigations in the present paper are of a purely theoretical character. Nevertheless they throw some light on the question of numerical

* Received October 30, 1947.

[1] For the notion of "fortsetzbare Anfangsbedingungen" see [7].

computation of solutions of hyperbolic equations. Such questions have recently attracted considerable attention among workers in the field of gas dynamics, in which hyperbolic differential equations play a prominent rôle. But even the questions of pure existence and uniqueness of certain hyperbolic differential equation problems are of considerable importance in the field of gas dynamics, see, e. g. [14]. The results of the present paper may be helpful in answering such questions.[2]

We have confined ourselves to considering systems of quasi-linear equations of first order. This is no serious restriction; for, it is well known that the initial problem for equations of higher than first order can always be reduced to the initial problem for a system of first order and that the initial problem for general nonlinear equations can be reduced by differentiation to a problem for quasi-linear ones, see, e. g. [13, Vol. II, § 7. 2]. Also the restriction to taking the x-axis as initial curve is not serious since it can be removed by a coordinate transformation.

For the general hyperbolic equation of second order for one unknown function the existence of a solution of the initial problem was first established by H. Lewy [1] in 1927 through reduction of the differential equation to a system of equations in characteristic form. In this form the equations are amenable to treatment by finite differences or iterations. The same procedure can also be used for systems of two equations of first order for two functions.

In 1928 H. Lewy and the present author published a paper [2] about hyperbolic differential equations of higher order for one unknown function. In the first part of this paper we reduced the equation to a system of equations in characteristic form; in the second part we treated a certain type of differential equations given in characteristic form. Unfortunately we overlooked the fact that the characteristic equations derived in the first part are not in general of the type considered in the second part.[3] Therefore the conclusions stated in the paper are not justified.

The difficulty met when one reduces the quasi-linear equation to characteristic form arises from the fact that only two of the N characteristic directions can be selected as new coordinate directions and that the equations involving other characteristic directions become quadratic in the derivatives.

[2] The investigations of the present paper were carried out in connection with work on gas dynamics done under contract No. N6ori-201, Task Order No. 1, with the Office of Naval Research.

[3] We owe this observation to C. DePrima. After having written this paper, the author became aware of the fact that M. Cinquini-Cibrario [15] has already made this observation and at the same time has given a new proof for the theorems in question. Her methods are quite different from those of the present paper.

in the two selected characteristic directions. In a sense one may say that the characteristic curves are rather sensitive to changes in the data of the problem [4] and, therefore, perhaps not so suitable as a basic working tool. This fact makes one wonder whether one can expect a good numerical approximation if such a reduction to characteristic form is used, together e. g. with the method of finite differences.[5]

In the present paper the method of reducing the quasi-linear differential equations to characteristic form is abandoned. This reduction is, however, used for linear equations. The solutions of quasi-linear and semi-linear equations are obtained from solutions of linear equations by iteration processes. In order to prove that these iteration processes converge and yield sufficiently strong results it is necessary to establish sufficiently strong properties of the solution of linear equations. Our treatment is in several respects related to that of Perron [3], who has proved existence and uniqueness theorems for linear and semi-linear hyperbolic equations; his assumptions, however, are not weak enough to make the transition from linear to quasi-linear equations possible.

Theorems on unique existence of solutions of the type considered here could be derived from the theorems concerning hyperbolic differential equations for functions of several variables proved by Schauder [8], Frankl [9], Petrowsky [10], Christianovitch [11]. When the number of independent variables is specialized to be two, the theorems of these authors reduce to statements about unique existence in which stronger differentiability conditions are required than in the theorem proved in the present paper. Aside from this fact, however, it seems justified to give an independent treatment of equations for functions of two independent variables because of the considerable intrinsic interest such equations have.

1. Basic notions. The differential equation which we shall investigate refers to a system $u = \{u^n\}$ of N functions $u_1, \cdots, u_N$, simply called "a function," of the two variables x and y. The coefficients of the equation, given as square matrices

$$a = \{a^{mn}\}, \qquad b = \{b^{mn}\}, \qquad\qquad m, n = 1, \cdots, N,$$

[4] Difficulties due to the sensitivity of the characteristics of one equation of first order were emphasized by Haar [4, 5, 6], see also [12].

[5] A recent different approach by R. Courant to the problem of hyperbolic equations, indicated in [14, II, 81] and to be published later in detail, leads to less pessimistic conclusions.

and the right member $g = \{g^n\}$ depend on x, y, and u. The equation can then be written in the form

$$\Lambda \cdot u = au_x + bu_y = g; \tag{1.1}$$

subscripts x and y, here and in the following, signify partial differentiation with respect to x and y. The variables x and y are restricted to domains $\mathcal{R}$ of the type characterized by

$$\mathcal{R}: X_1 \leq x \leq X_2, \qquad 0 \leq y \leq Y(x),$$

$Y(x)$ being an appropriate non-negative continuous function defined for $X_1 \leq x \leq X_2$. The value $[u = \bar{u}(x)]$ of the function u is prescribed on the initial segment

$$\mathcal{I}: X_1 \leq x \leq X_2, \qquad y = 0,$$

cut out from $\mathcal{R}$ by the axis $y = 0$.

To express restrictions for the function u we introduce the "absolute value"

$$|u(x, y)| = \max_n |u^n(x, y)|. \tag{1.2}$$

Then, with a given positive constant Ω, we shall frequently restrict u by the condition

$$|u(x, y)| \leq \Omega.$$

The domain of values x, y, and u, given by the condition that (x, y) be in $\mathcal{R}$ and $|u| \leq \Omega$ will be denoted by

$$\mathcal{R}_\Omega: \quad (x, y) \text{ in } \mathcal{R}, \ |u| \leq \Omega.$$

The matrixes a and b, and the vector g are then supposed to be defined in such a domain $\mathcal{R}_\Omega$ as continuous functions of x, y, and u. In addition, the matrices a and b are to have continuous derivatives.

The decisive property we require of the matrices a and b is that for each x, y, u in $\mathcal{R}_\Omega$ the elementary divisors of the matrix $a - \kappa b$, in which κ is a parameter, are simple and that the eigenvalues $k^1, \cdots, k^N$ are real. More specifically, we require that there exist in $\mathcal{R}_\Omega$ matrices p and q with non-vanishing determinant such that

$$p\,b\,q = 1 = \begin{pmatrix} 1 & 0 \cdots 0 \\ 0 & 1 \cdots 0 \\ \cdot & \cdots\cdots \\ \cdot & \cdots\cdots \\ 0 & 0 \cdots 1 \end{pmatrix}, \tag{1.3}$$

$$p\,a\,q = k = \begin{pmatrix} k^1 & 0 \cdots 0 \\ 0 & k^2 \cdots 0 \\ \cdot & \cdots\cdots \\ \cdot & \cdots\cdots \\ \cdot & \cdots\cdots \\ 0 & 0 \cdots k^N \end{pmatrix}. \tag{1.4}$$

The differential equation (1.1) is called "hyperbolic" if this condition is satisfied.

The significance of the eigenvalues k^n is that the N directions in the (x, y)-plane given by

$$dx = k^n dy, \qquad n = 1, \cdots, N, \tag{1.5}$$

are characteristic directions for the differential equation. These directions depend, of course, on the point (x, y) and on the value of u assigned to it. The fact, implied by our assumptions, that all k^n are finite, expresses that a line $y = \text{const.}$ never has a characteristic direction.

The matrices p and q depend on x, y, and u and it is, in addition, assumed that these matrices p and q can be so found that they possess continuous derivatives with respect to x, y, and u in $\mathcal{R}_\Omega$.

The requirements formulated so far will in the following always be imposed tacitly; they will in general not be mentioned explicitly any more.

2. Notations. It has been found useful to employ in the following quite a number of notations which we shall explain and compile in this section. For vectors $f = \{f^n\}$ and matrices $\{c^{mn}\}$ we introduce, in agreement with (1.2), the absolute value

$$|f| = \max_n |f^n|, \qquad \max_n = \max_{n=1,\ldots,N}, \tag{2.01}$$

$$|c| = \max_m \sum_n |c^{mn}|, \qquad \max_m = \max_{n=1,\ldots,N}, \sum_n = \sum_{n=1}. \tag{2.02}$$

Clearly,

$$|cf| \leq |c|\ |f|. \tag{2.03}$$

If f or c depends on x, y, u in $\mathcal{R}_\Omega$ we introduce the "norm"

$$\| f \| = \max_{\mathcal{R}_\Omega} |f(x, y, u)|, \qquad \max_{\mathcal{R}_\Omega} = \max_{x, y, u \text{ in } \mathcal{R}_\Omega}, \tag{2.04}$$

$$\| c \| = \max_{\mathcal{R}_\Omega} |c(x, y, u)|. \tag{2.05}$$

Clearly,

$$\| cf \| \leq \| c \|\ \| f \|. \tag{2.06}$$

On occasion we make use of the fact that the norm of f could also be characterized by

(2.07) $$\| f \| = \max_n \max_{\mathcal{R}_\Omega} | f^n(x, y, u) | .$$

If f, or c, does not depend on u, the norm $\| f \|$, or $\| c \|$, is simply defined with reference to the domain $\mathcal{R}$ instead of the domain $\mathcal{R}_\Omega$.

On occasion we shall find it convenient to employ a still simpler notation for the norms, setting

(2.08) $$\| f \| = F,$$

and generally denoting by a capital Latin letter the norm of a quantity denoted by the corresponding small Latin letter.

The pair of derivatives of a quantity f, with respect to x and y, will be denoted by

(2.09) $$f_{x,y};$$

its norm by

(2.10) $$\| f_{x,y} \| = \max\{\| f_x \|, \| f_y \|\} = F_1.$$

The matrix $\{f^m{}_{u^n}\}$ of derivatives of a quantity $f = \{f^m\}$ with respect to the variables u^n will be denoted by

(2.11) $$f_u = \{f^m{}_{u^n}\}.$$

Its norm will be denoted by

(2.12) $$\| f_u \| = F_u.$$

If f and g are vectors we introduce the vector $f_u g$ by

(2.13) $$f_u g = \{\sum_n f^m{}_{u^n} g^n\}.$$

All these stipulations refer in obvious manner to matrices just as to vectors.

From the operator $\wedge$ which transforms a function u' into

(2.14) $$\wedge \cdot u' = a u'_x + b u'_y,$$

see (1.1), we derive by differentiation the operator $\wedge_{x,y} + (\wedge_u u_{x,y})$ which transforms the function u' into

(2.15) $$\wedge_{x,y} \cdot u' + (\wedge_u u_{x,y}) \cdot u'$$
$$= \{a_x u'_x + b_x u'_y + a_u u_x u'_x + b_u u_x u'_y,\ a_y u'_x + b_y u'_y + a_u u_y u'_x + b_u u_y u'_y\}.$$

For the pair of matrices a and b we introduce the norm

$$\| a \| + \| b \| = A. \tag{2.16}$$

For the derivatives of a and b we set

$$\| a_{x,y} \| + \| b_{x,y} \| = A_1, \qquad \| a_u \| + \| b_u \| = A_u. \tag{2.17}$$

Clearly, we have by (2.10)

$$\| \wedge \cdot u' \| \leq A \, \| u'_{x,y} \| = A U'_1, \tag{2.18}$$

$$\| \wedge_{x,y} \cdot u' + (\wedge_u u_{x,y}) \cdot u' \| \leq (A_1 + A_u U_1) U'_1. \tag{2.19}$$

For the initial function we shall frequently use the expression

$$\| \bar{u} - \bar{\bar{u}} \| = \max_{\mathfrak{I}} | u(\bar{x}) - u(\bar{\bar{x}}) | , \qquad \max_{\mathfrak{I}} = \max_{\bar{x} \text{ in } \mathfrak{I}, \, \bar{\bar{x}} \text{ in } \mathfrak{I}} . \tag{2.20}$$

For the inverse matrix q^{-1} we shall employ the expression

$$\| q^{-1} - \bar{q}^{-1} \| = \max_{\mathcal{R}_\Omega, \mathfrak{I}} | q^{-1}(x, y, u) - q^{-1}(\bar{x}, 0, \bar{u}(\bar{x})) | , \tag{2.21}$$

$$\max_{\mathcal{R}_\Omega, \mathfrak{I}} = \max_{x, y, u \text{ in } \mathcal{R}_\Omega, \, \bar{x} \text{ in } \mathfrak{I}} ,$$

and use the notation

$$L_0 = \| q \| \; \| q^{-1} - \bar{q}^{-1} \| \tag{2.22}$$

or more generally

$$L_\lambda = \| q \| \; \| q^{-1} - \bar{q}^{-1} e^{-\lambda y} \| . \tag{2.23}$$

Further we set

$$J = \| q^{-1} \| \; \| q \| . \tag{2.24}$$

Incidentally, the inverse q^{-1} of the matrix q will also be denoted by

$$\hat{q} = q^{-1}. \tag{2.25}$$

and we set accordingly

$$\| q^{-1} \| = \hat{Q}. \tag{2.26}$$

Finally, we note that all notations introduced with reference to first derivatives will be employed for second derivatives in an obviously analogous manner.

3. Reduction to characteristic form. We first consider the case in which the matrices a and b do not depend on u. The same is then true for p, q, and k. We introduce the N families of characteristic lines

$$\mathcal{C}^m : \; x = x^m(t, \xi), \qquad y = t, \qquad\qquad m = 1, \cdots, N,$$

the functions $x^m(t,\xi)$ being the solutions of the differential equations

(3.01) $$x_t = k^m(x,t)$$

with the initial condition

(3.02) $$x = \xi \text{ for } t = 0,\ \xi \text{ on } \mathfrak{J}.$$

Since the function $k^m(x,t)$ defined by (1.4), was assumed to possess continuous derivatives for (x,t) in $\mathfrak{R}$, through every point $x = \xi$ on $\mathfrak{J}$ such characteristic lines can be passed and continued up to the boundary of the domain $\mathfrak{R}$. Moreover, the function $x^m(t,\xi)$ possesses the continuous derivative

(3.03) $$\zeta^m = x^m{}_\xi(t,\xi)$$

with respect to the parameter ξ, satisfying the differential equation

(3.04) $$\zeta_t{}^m = k_x{}^m \zeta^m$$

and the initial conditions

(3.05) $$\zeta^m = 1 \text{ for } t = 0.$$

It then follows that

(3.06) $$\zeta^m \neq 0 \text{ in } \mathfrak{R}, \text{ wherever defined.}$$

We must restrict ourselves to such domains $\mathfrak{R}$ which have the property that the N characteristic lines through any point in $\mathfrak{R}$ can be continued up to the initial segment $\mathfrak{J}$ or, in other words, that every point in $\mathfrak{R}$ can be reached from the initial section by N characteristic lines. Such domains $\mathfrak{R}$ we shall refer to as "domains of determinacy." It evidently depends on the matrices a, b or rather on the diagonal matrix k, as functions of x, y in $\mathfrak{R}$, whether or not $\mathfrak{R}$ is a domain of determinacy. Later on we shall make use of the following

Remark. Let $\mathfrak{R}$ be a determinacy domain. Let the point x_0 lie on the initial segment $\mathfrak{J}$ and the number ϵ be such that the interval $\mathfrak{J}'$: $|x - x_0| \leq \epsilon$ lies in $\mathfrak{J}$. Let Y be a positive number. Then the domain

$$\mathfrak{R}': \ |x - x_0| + Ky \leq \epsilon, \qquad 0 \leq y \leq Y, \qquad x, y \text{ in } \mathfrak{R}$$

is a determinacy domain.

Here $K = \| k \| = \max\limits_n \| k^n \|$ is the norm of the diagonal matrix k. Consider a boundary point $\mathfrak{P}$ of $\mathfrak{R}'$ with $y > 0$ and the N characteristic directions through it with decreasing y. If $\mathfrak{P}$ is a boundary point of $\mathfrak{R}$ the N directions lead into $\mathfrak{R}$. If $\mathfrak{P}$ is a boundary point of the triangle $|x - x_0| + Ky \leq \epsilon$, the N directions lead into the triangle because of

$|dx/dy| = |k^m| \leq K$. *If* $\mathfrak{P}$ lies on $y = Y$, the N directions lead to $y \leq Y$. As a result the N directions lead to the region common to $\mathfrak{R}$, the triangle, and $y \leq Y$; i.e. to the domain $\mathfrak{R}'$. Any characteristic line through any point of $\mathfrak{R}'$ when continued to the boundary of $\mathfrak{R}'$, therefore, ends up on $\mathfrak{I}'$.

To clarify the notion of domain of determinacy we contrast it with the notion of "domain of dependence" of a point (x, y) in $\mathfrak{R}$, which is the common part of all domains of determinacy containing this point.[6] It is clear that this domain of dependence is also a determinacy domain. Therefore *the uniqueness theorems* which we shall prove will *also apply to the domain of dependence of a point.* It will, therefore, follow that the value of the solution at this point is not affected if the data of the problem, i.e. the coefficients of the differential equation and the initial function, are modified outside of the domain of dependence.

In the following we shall always *assume that* $\mathfrak{R}$ *is a domain of determinacy.* As a consequence the functions $x = x^m(y, \xi)$ possess inverse functions

$$\xi = \xi^m(x, y) \tag{3.07}$$

defined for every point (x, y) in $\mathfrak{R}$. By virtue of (3.06) the function $\xi^m(x, y)$ possesses continuous derivatives with respect to x and y lies in $\mathfrak{I}$. The pair $t = y$, $\xi = \xi^m(x, y)$ maps the domain $\mathfrak{R}$ on a domain

$$P^m\colon\ 0 \leq t \leq T^m(\xi), \qquad \xi \text{ in } \mathfrak{I},$$

of the (t, ξ)-plane, $T^m(\xi)$ being the value of $t > 0$ for which the characteristic $x = x^m(t, \xi)$, $y = t$ meets the boundary of $\mathfrak{R}$.

We now transform the differential equation $\Lambda \cdot u = au_x + bu_y = g$ into characteristic form. To this end we introduce new functions $v = \{v^n\}$ by

$$u = qv, \tag{3.08}$$

the matrix q having been introduced in **1** in connection with (1.3) and (1.4). Applying the matrix p to $\Lambda \cdot u = g$ we obtain the characteristic form of the differential equation

$$kv_x + v_y = h, \tag{3.09}$$

in which $h = \{h^n\}$ is given by

$$h = p(g - aq_xv - bq_yv). \tag{3.10}$$

---

[6] If all eigen-values k^n are different, the domain of dependence of a point is bounded by two "outer" characteristics through this point. If some of the eigen-values are equal at some places, this boundary may consist of sections of different characteristics.

This characteristic form of the equation can now easily be referred to characteristic parameters t. Introducing t and ξ as new variables by $x = x^m(t, \xi)$, $y = t$, equations (3.09) take immediately the form

$$(3.11) \qquad v_t{}^n = h^n, \qquad n = 1, \cdots, N,$$

for the functions $v^n = v^n(x^n(t, \xi), t)$ defined in P^n.

4. Linear equations in characteristic form. In this section we consider linear differential equations which from the outset are given in characteristic form (3.09) without assuming that this form was obtained by a transformation (3.08) from the general form (1.1). Accordingly, we assume that a diagonal matrix $k = \{k^n\}$ is given as a continuously differentiable function in a region $\mathcal{R}$, which is, of course, assumed to be a determinacy region. Further we assume that a matrix r and a vector s are defined as continuous functions in $\mathcal{R}$. Finally a continuous function $\bar{v}(x)$ should be defined on the initial section $\mathcal{J}$ of $\mathcal{R}$.

For functions $v = \{v^n\}$ of x and y in $\mathcal{R}$ we then consider the differential equation

$$(4.01) \qquad kv_x + v_y + rv = s \text{ in } \mathcal{R}$$

and the initial condition

$$(4.02) \qquad v = \bar{v} \text{ in } \mathcal{J}.$$

We introduce the new variables t and ξ through the functions

$$(4.03) \qquad x = x^m(t, \xi), \qquad y = t,$$

see Section **3**, which have continuous derivatives with respect to x and y in P^m and satisfy the equation (3.01) $x_t{}^m = k^m(x^m, t)$.

Let $v^n(x, y)$, $n = 1, \cdots, N$, be a set of N continuous functions in $\mathcal{R}$ such that the functions

$$v^n(x^n(t, \xi), t)$$

of t and ξ in P^n possess continuous derivatives with respect to t and satisfy the equations

$$(4.04) \qquad v_t{}^n + (rv)^n = s^n.$$

Then we say that the function $v = \{v^n\}$ satisfies equation (4.01) $kv_x + v_y + rv = s$ *in the wider sense.* If, in addition, the function v possesses continuous derivatives v_x, v_y, it satisfies this equation also in *the strict sense.*

Our basic tool will be estimates of the norm of the solution v of (4.04) in terms of the coefficients and the initial function.

LEMMA 4.1. *Suppose the function v in $\mathcal{R}$ satisfies equation (4.01) in the wider sense and the initial condition $v = \bar{v}$ on $\mathcal{I}$. Then for any positive number λ the estimate*

$$(4.05)\qquad [\lambda - R]\, \| ve^{-\lambda y} \| \leq \lambda \, \| \bar{v} \| + \| se^{-\lambda y} \|$$

holds, with $R = \| r \|$, see (2.08).

Proof. We observe that each component $v^n(x^n(t,\xi),t)$ satisfies the equation $v_t^n = -(rv)^n + s^n$; hence the inequality

$$| v_t^n | \leq | (rv)^n | + | s^n | \leq \sum_m | r^{nm} | \max_m | v^m | + | s^n |$$

is valid. After multiplying the right hand side by $e^{\lambda t}e^{-\lambda y} = 1$ we find

$$| v_t^n | \leq e^{\lambda t} [\max_{\mathcal{R}} \sum_m | r^{nm} | \max_{\mathcal{R}} \max_m | v^m e^{-\lambda y} | + \max_{\mathcal{R}} | s^n e^{-\lambda y} |];$$

hence, by integration,

$$\lambda \, | v^n(x^n(t,\xi),t) - v^n(\xi,0) |$$
$$\leq e^{\lambda t} [\max_{\mathcal{R}} \sum_n | r^{nm} | \max_{\mathcal{R}} \max_m | v^m e^{-\lambda y} | + \max_{\mathcal{R}} | s^n e^{-\lambda y} |].$$

Using the notations (2.01), (2.04), (2.07), (2.02), (2.05) the last inequality can be written in the form

$$\lambda \, | v^n(x^n(t,\xi),t) - v^n(\xi,0) | \leq e^{\lambda t} [\| r \| \; \| ve^{-\lambda y} \| + \| se^{-\lambda y} \|]$$

or, using the inverse function $\xi^n(x,y)$ introduced in **3**, see (3.07),

$$(4.06)\qquad \lambda \, | v^n(x,y) - \bar{v}^n(\xi^n(x,y)) | \leq e^{\lambda y} [\| r \| \; \| ve^{-\lambda y} \| + \| se^{-\lambda y} \|].$$

From $| v^n(x,y) | \leq | v^n(x,y) - \bar{v}^n(\xi^n(x,y)) | + \| \bar{v} \|$ we then obtain

$$| v^n(x,y) | \leq \| \bar{v} \| + \lambda^{-1} e^{\lambda y} [\| r \| \; \| ve^{-\lambda y} \| + \| se^{-\lambda y} \|]$$

and, consequently, (4.05).

Estimate (4.05) is sufficient for many purposes, but in addition we need later on a more refined estimate. Let $q(x,y)$ be a matrix which, together with its inverse $q^{-1}(x,y)$, is a continuous function of x, y in $\mathcal{R}$. Only later on shall we identify q with the matrix introduced in **1**. Then we can derive an estimate which differs from (4.05) in that qv takes the place of v and the terms are multiplied by different factors. It is important for certain applications in the following that the ratio of the factors of $\| qve^{-\lambda y} \|$ and $\| \bar{q}\bar{v} \|$ approach the value 1 as λ increases indefinitely. This requirement accounts for the slight complications in the formula given below and its derivation.

LEMMA 4.2. *Suppose that the function v in $\mathcal{R}$ satisfies the equation* (4.01) *in the wider sense and the initial condition $v=\bar{v}$ on $\mathcal{J}$. Then, for any positive number λ and any matrix q with the properties just indicated, the estimate*

$$(4.07) \qquad [\lambda - JR] \, \| qve^{-\lambda y} \| \\ \leq \lambda(1+L_\lambda) \| \bar{q}\bar{v} \| + \lambda J \, \| \bar{\bar{q}}\bar{\bar{v}} - \bar{q}\bar{v} \| + Q \, \| se^{-\lambda y} \|$$

holds. Here the quantities J and L_λ are defined by (2.24) *and* (2.23). *The definition of $\| \bar{q}\bar{v} - \bar{\bar{q}}\bar{\bar{v}} \|$ corresponds to* (2.20).

Proof. We set $w(\bar{x}; x, y) = q^{-1}(x,y)\bar{q}(\bar{x})\bar{v}(\bar{x})$ with arbitrary $\bar{x}$ in $\mathcal{J}$, x, y in $\mathcal{R}$. Using the functions $\xi^n(x,y)$ introduced in **3**, see (3.07), we have

$$| v^n(x,y) - e^{\lambda y} w^n(\bar{x}; x, y) | \leq | e^{\lambda y} w^n(\bar{x}; x, y) - \bar{v}^n(\xi^n(x,y)) | \\ + | v^n(x,y) - \bar{v}^n(\xi^n(x,y)) | .$$

From this relation and (4.06) we deduce

$$\lambda \, | e^{-\lambda y} v^n(x,y) - w^n(\bar{x}; x, y) | \\ \leq \lambda \, | q^{-1}(x,y)\bar{q}(\bar{x})\bar{v}(\bar{x}) - e^{-\lambda y}\bar{v}(\xi^n(x,y)) | + \| r \| \; \| ve^{-\lambda y} \| + \| se^{-\lambda y} \| ,$$

whence

$$\lambda \, | e^{-\lambda y} v^n(x,y) - w^n(\bar{x}; x, y) | \leq \lambda \, \| q^{-1} \| \; | \bar{q}(\bar{x})\bar{v}(\bar{x}) - \bar{q}(\xi^n(x,y))\bar{v}(\xi^n(x,y)) | \\ + \lambda \, | q^{-1}(x,y) - e^{-\lambda y}\bar{q}^{-1}(\xi^n(x,y)) | \; | \bar{q}(\xi^n(x,y))\bar{v}(\xi^n(x,y)) | \\ + \| r \| \; \| q^{-1} \| \; \| qve^{-\lambda y} \| + \| se^{-\lambda y} \| .$$

Since $\xi^n(x,y)$ lies in the interval $\mathcal{J}$ we obtain

$$\lambda \, | q^{-1}(x,y) \{ e^{-\lambda y} q(x,y) v(x,y) - \bar{q}(\bar{x})\bar{v}(\bar{x}) \} | \\ = \lambda \, | e^{-\lambda y} v(x,y) - q^{-1}(x,y)\bar{q}(\bar{x})\bar{v}(\bar{x}) | \\ \leq \lambda \, \| q^{-1} \| \; \| \bar{q}\bar{v} - \overline{\bar{q}\bar{v}} \| + \lambda \, \| q^{-1} - e^{-\lambda y}\bar{q}^{-1} \| \; \| \bar{q}\bar{v} \| \\ + \| r \| \; \| q^{-1} \| \; \| qve^{-\lambda y} \| + \| se^{-\lambda y} \| ,$$

whence

$$\lambda \, \| e^{-\lambda y} qv - \overline{\bar{q}\bar{v}} \| \leq \lambda \, \| q \| \; \| q^{-1} \| \; \| \bar{q}\bar{v} - \bar{\bar{q}}\bar{\bar{v}} \| \\ + \lambda \, \| q \| \; \| q^{-1} - e^{-\lambda y}\bar{q}^{-1} \| \; \| \bar{q}\bar{v} \| \\ + \| q \| \; \| q^{-1} \| \; \| r \| \; \| qve^{-\lambda y} \| + \| q \| \; \| se^{-\lambda y} \|$$

and, consequently, (4.07).

An immediate consequence of Lemma 4.1 is the uniqueness

THEOREM 4.1. *There exists in $\mathcal{R}$ at most one solution v of equation*

(4.01), $kv_x + v_y + rv = s$, *in the wider sense assuming on* $\mathfrak{J}$ *the given initial value* $v = \overline{v}$.

We need only use relation (4.05) with $\overline{v} = 0$, $s = 0$, and $\lambda > R$ for the difference of two solutions of (4.01). Another immediate consequence of Lemma 4.1 is

LEMMA 4.3. *Let* $r^{(\sigma)}$, $s^{(\sigma)}$ *be sequences of continuous functions in* $\mathfrak{R}$ *such that the norms* $\| r^{(\sigma)} \| = R^{(\sigma)}$ *remain bounded,* $R^{(\sigma)} \leq \rho$, *while* $\| s^{(\sigma)} \|$ *approaches zero. Let* $v^{(\sigma)}$ *be a solution in the wider sense of*

$$kv_x^{(\sigma)} + v_y^{(\sigma)} + r^{(\sigma)} v^{(\sigma)} = s^{(\sigma)} \tag{4.08}$$

assuming the initial value $v^{(\sigma)} = 0$ *on* $\mathfrak{J}$. *Then* $\| v^{(\sigma)} \| \to 0$.

We need only choose $\lambda > \rho$; then (4.05) yields

$$[\lambda - \rho] \, \| v^{(\sigma)} e^{-\lambda y} \| \leq \| s^{(\sigma)} e^{-\lambda y} \| \to 0$$

and hence the statement

LEMMA 4.4. *Let* $r^{(\sigma)}$, $s^{(\sigma)}$ *be sequences of continuous functions in* $\mathfrak{R}$ *which converge uniformly in* $\mathfrak{R}$ *to continuous limit functions* r, s: $\| r^{(\sigma)} - r \| \to 0$, $\| s^{(\sigma)} - s \| \to 0$ *as* $\sigma \to \infty$. *Let* $v^{(\sigma)}$ *be a solution in the wider sense of the equation* (4.08) *assuming the initial value* $v^{(\sigma)} = \overline{v}$ *on* $\mathfrak{J}$, *independently of* σ, *and which converges uniformly in* $\mathfrak{R}$ *to a continuous function* v. *Then* v *satisfies equation* (4.01) *in the wider sense. Incidentally, such a sequence* $v^{(\sigma)}$ *always does converge uniformly in* $\mathfrak{R}$ *to a continuous limit function.*

Indeed, by virtue of the transformation $x = x^n(\xi, t)$, $y = t$, each component $v^{(\sigma)n}$ is a function of t and ξ in P^n which satisfies the integral equation

$$v^{(\sigma)n}(t, \xi) = \overline{v}^n(\xi) + \int_0^t (- r^{(\sigma)} v^{(\sigma)} + s^{(\sigma)})^n (t', \xi) \, dt'.$$

Hence the limit function $v^n(t, \xi)$ satisfies

$$v^n(t, \xi) = \overline{v}^n(\xi) + \int_0^t (- rv + s)^n (t', \xi) \, dt',$$

whence the statement follows.

To show that such a sequence $v^{(\sigma)}$ always has a limit function as indicated we choose a number ρ such that $\| r^{(\sigma)} \| \leq \rho$ and a number $\lambda > \rho$. Inequality (4.05) then shows that $\| v^{(\sigma)} \|$ is bounded. Next we introduce for any σ and τ the differences

$$v^{(\sigma\tau)} = v^{(\sigma)} - v^{(\tau)}, \quad r^{(\sigma\tau)} = r^{(\sigma)} - r^{(\tau)}, \quad s^{(\sigma\tau)} = s^{(\sigma)} - s^{(\tau)}$$

and observe that $v^{(\sigma\tau)}$ satisfies in the wider sense the equation

$$kv_x^{(\sigma\tau)} + v_y^{(\sigma\tau)} + r^{(\sigma)}v^{(\sigma\tau)} = -r^{(\sigma\tau)}v^{(\tau)} + s^{(\sigma\tau)}$$

and assumes the initial value zero. Lemma 4.3 then may be applied; it yields $\| v^{(\sigma\tau)} \| \to 0$. Consequently, $v^{(\sigma)}$ converges uniformly in $\mathcal{R}$ to a continuous limit function.

The existence theorem of this section is

THEOREM 4.2. *There exists in $\mathcal{R}$ a solution of the equation*

$$kv_x + v_y + rv = s \tag{4.01}$$

in the wider sense, assuming on $\mathcal{J}$ the given initial value $v = \bar{v}$.

Proof. We set up iterations. We construct a sequence of continuous function $v^{(\sigma)}$ in $\mathcal{R}$ assuming the initial value $\bar{v}$, beginning with $v^{(0)}(x,y) = \bar{v}(x)$. Having found a continuous function $v^{(\sigma)}(x,y)$ we shall determine $v^{(\sigma+1)}$ as a solution of

$$kv_x^{(\sigma+1)} + v_y^{(\sigma+1)} = -rv^{(\sigma)} + s. \tag{4.09}$$

The right member here is a continuous function of x and y in $\mathcal{R}$ and becomes a continuous function of t and ξ in P^n after the transformation $x = x^n(t,\xi)$, $y = t$. We now determine the component $v^{(\sigma+1)n}$ as function of t and ξ in P^n through integration so that it satisfies

$$v_t^{(\sigma+1)n} = (-rv^{(\sigma)} + s)^n \text{ in } P^n, \quad v^{(\sigma+1)n} = \bar{v}^n \text{ for } t = 0. \tag{4.10}$$

The function $v^{(\sigma+1)n}$ so determined depends continuously on t and ξ in P^n, as is well known, since the right member does. Now, by virtue of the inverse transformation (3.07) $\xi = \xi^n(x,y)$, $t = y$, the component $v^{(\sigma+1)n}$ becomes a continuous function of x and y in $\mathcal{R}$. The function $v^{(\sigma+1)} = \{v^{(\sigma+1)n}\}$, therefore, is continuous in $\mathcal{R}$, assumes the initial value $v^{(\sigma+1)} = \bar{v}$ and satisfies in the wider sense the equation (4.09). The iterations can therefore be carried out.

The differences $\Delta v^{(\sigma+1)} = v^{(\sigma+1)} - v^{(\sigma)}$ satisfy the equations

$$k\Delta v_x^{(\sigma+1)} + \Delta v_y^{(\sigma+1)} = -r\Delta v^{(\sigma)}, \qquad \sigma = 2, 3, \cdots \tag{4.11}$$

in the wider sense and assume the initial values $\Delta v^{(\sigma+1)} = 0$ on $\mathcal{J}$. Applying Lemma 4.1 with 0 instead of r, and $-r\Delta v^{(\sigma)}$ instead of s, we obtain from (4.05) the estimate

$$\lambda \| \Delta v^{(\sigma+1)} e^{-\lambda y} \| \leq R \| \Delta v^{(\sigma)} e^{-\lambda y} \| . \tag{4.12}$$

We choose a positive number $\theta < 1$, and a positive number λ so large that

$$R = \| r \| \leq \theta\lambda. \tag{4.13}$$

Then (4.12) gives

$$\| \Delta v^{(\sigma+1)} e^{-\lambda y} \| \leq \theta \, \| \Delta v^{(\sigma)} e^{-\lambda y} \| ,$$

whence by induction

$$\| \Delta v^{(\sigma)} e^{-\lambda y} \| \leq \theta^{\sigma-1} \, \| \Delta v^{(1)} e^{-\lambda y} \| .$$

The uniform convergence of $v^{(\sigma)} e^{-\lambda y}$ and, hence, of $v^{(\sigma)}$ to a continuous limit function then follows. That this limit function satisfies equation (4.01) in the wider sense follows from Lemma 4.4, applied for $r^{(\sigma)} = 0$ and $s^{(\sigma)} = - r v^{(\sigma)} + s$.

We now formulate the first differentiability theorem,

THEOREM 4.3. *Suppose that not only the diagonal matrix k, but in addition the matrix r and the vector s, have continuous derivatives with respect to x and y in $\mathcal{R}$, and that $\bar{v}$ has a continuous derivative with respect to x in $\mathfrak{J}$. Then a solution of equation (4.01) in the wider sense assuming the initial value $\bar{v}$ possesses continuous derivatives with respect to x and y in $\mathcal{R}$. The function v satisfies equation (4.01) in the strict sense. The derivatives v_x and v_y satisfy in the wider sense the equation*

$$k(v_{x,y})_x + (v_{x,y})_y + k_{x,y} v_x + r v_{x,y} + r_{x,y} v = s_{x,y}, \tag{4.14}$$

see (2.09), obtained by formal differentiation of equation (4.01). The initial value $\overline{v_x}$ of v_x is $\overline{v_x} = \bar{v}_x$ while the initial value $\overline{v_y}$ of v_y is to be determined from relation (4.01).

The proof of Theorem 4.3 will be given as a corollary to the existence Theorem 4.2. It would be possible to give an independent proof, but the present procedure, in which we follow Perron [3], appears to be simpler.

Consider the functions $v^{(\sigma)}$ as defined through iterations in the proof of Theorem 4.2. We maintain that they have continuous derivatives with respect to x and y in $\mathcal{R}$ under the assumptions of Theorem 4.3. This is certainly the case for $v^{(0)} = \bar{v}(x)$. Suppose it were true for $v^{(\sigma)}$. Then the function $-(r v^{(\sigma)} + s)^n$, considered a function of t and ξ in P^n after the transformation $x = x^n(t, \xi)$, $y = t$, possesses continuous derivatives with respect to t and ξ since this is the case for $x^n(t, \xi)$. The same is then true for the function $v^{(\sigma+1)n}$ of t and ξ determined from (4.10). Returning to the variables x and y by the transformation $\xi = \xi^n(y,x)$, $t = y$, we see that $v^{(\sigma+1)}$ has continuous derivatives with respect to x and y in $\mathcal{R}$. Thus it follows by induction that $v^{(\sigma)}$ has continuous derivatives for all σ.

From relation (4.10) it follows that the derivatives of $v^{(\sigma+1)}$ with respect to t and ξ satisfy the equations

$$(v_t{}^{(\sigma+1)n})_t = (-rv^{(\sigma)} + s)_t{}^n, \quad (v^{(\sigma+1)n}{}_\xi)_t = (-rv^{(\sigma)} + s)^n{}_\xi.$$

The derivatives

$$v_x{}^{(\sigma+1)n} = v^{(\sigma+1)n}{}_\xi/\zeta^n, \quad v_y{}^{(\sigma+1)n} = v_t{}^{(\sigma+1)n} + k^n v^{(\sigma+1)n}{}_\xi/\zeta^n$$

therefore satisfy, in view of (3.04), the equations

$$(v_x{}^{(\sigma+1)n})_t + k_x{}^n v_x{}^{(\sigma+1)n} = (-rv^{(\sigma)} + s)_x{}^n,$$
$$(v_y{}^{(\sigma+1)n})_t + k_y{}^n v_x{}^{(\sigma+1)n} = (-rv^{(\sigma)} + s)_y{}^n.$$

In other words, $v_{x,y}{}^{(\sigma+1)}$ satisfies equation

$$k(v_{x,y}{}^{(\sigma+1)})_x + (v_{x,y}{}^{(\sigma+1)})_y + k_{x,y}v_x{}^{(\sigma+1)} = (-rv^{(\sigma)} + s)_{x,y}$$

in the wider sense. The differences $\Delta v_{x,y}{}^{(\sigma+1)} = v_{x,y}{}^{(\sigma+1)} - v_{x,y}{}^{(\sigma)}$ satisfy the equations

$$k(\Delta v_{x,y}{}^{(\sigma+1)})_x + (\Delta v_{x,y}{}^{(\sigma+1)})_y + k_{x,y}v_x{}^{(\sigma+1)} = -(r\Delta v^{(\sigma)})_{x,y}$$

and assume the initial values zero. Hence Lemma 4.1 with $k_{x,y}$ instead of r and $-(r\Delta v^{(\sigma)})_{x,y}$ instead of s can be applied to the pair $\Delta v_{x,y}{}^{(\sigma+1)}$ (which embodies $2N$ instead of N functions); inequality (4.05) yields

$$(4.15) \qquad [\lambda - K_1] \, \| \Delta v_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| \leq R \, \| \Delta v_{x,y}{}^{(\sigma)} e^{-\lambda y} \| + R_1 \, \| \Delta v^{(\sigma)} e^{-\lambda y} \|$$

with $K_1 = \| k_{x,y} \|$, $R = \| r \|$, $R_1 = \| r_{x,y} \|$. If now λ is so chosen that

$$(4.16) \qquad R + R_1 \leq \theta(\lambda - K_1),$$

cf. (4.13), we find by combining (4.15) with (4.12)

$$\| \Delta v_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| + \| \Delta v^{(\sigma+1)} e^{-\lambda y} \| \leq \theta(\| \Delta v_{x,y}{}^{(\sigma)} e^{-\lambda y} \| + \| \Delta v^{(\sigma)} e^{-\lambda y} \|).$$

It is clear that $v_{x,y}{}^{(\sigma)}$ together with $v^{(\sigma)}$ converge to continuous limit functions $v_{x,y}{}^{\infty}$ and v^{∞}. It is also clear that $v_{x,y}{}^{\infty}$ are the derivatives of v^{∞}. It follows from Lemma 4.3 that v^{∞} and $v_{x,y}{}^{\infty}$ satisfy equations (4.01) and (4.11) in the wider sense. Since v^{∞} assumes the value $\bar{v}$ on $\mathfrak{I}$, Theorem 4.1 insures that $v^{\infty} = v$. Hence v possesses continuous derivatives v_x, v_y satisfying (4.11) while itself satisfies (4.01) in the strict sense. Thus Theorem 4.3 is proved.

For later purposes we need a differentiability theorem which is somewhat sharper than Theorem 4.3.

THEOREM 4.4. *Let $q(x, y)$ be a matrix*[7] *which together with its inverse $q^{-1} = \hat{q}(x, y)$ is a continuously differentiable function of x and y in $\mathfrak{R}$.*

---

[7] In later sections q will be identified with the transformation introduced in 1.

Suppose that the diagonal matrix k and in addition the matrices r^* and the vector s have continuous derivatives with respect to x and y in $\mathcal{R}$, and that $\bar{v}$ has a continuous derivative with respect to x in $\mathcal{J}$. Then a solution v of equation

$$kv_x + v_y + (k\hat{q}q_x + \hat{q}q_y)v + r^*v = s \tag{4.17}$$

in the wider sense assuming the initial value $\bar{v}$ possesses continuous derivatives $v_{x,y}$ with respect to x and y in $\mathcal{R}$. The function v satisfies equation (4.17) in the strict sense. This equation can therefore be written in the form

$$k\hat{q}(qv)_x + \hat{q}(qv)_y + r^*v = s.$$

The quantities $\hat{q}(qv)_{x,y}$ satisfy in the wider sense the equation

$$\begin{aligned} k(\hat{q}(qv)_{x,y})_x &+ (\hat{q}(qv)_{x,y})_y \\ &+ k_{x,y}\hat{q}(qv)_x + k(\hat{q}_{x,y})(qv)_x + (\hat{q}_{x,y})(qv)_y \\ &- (k\hat{q}_x + \hat{q}_y)(qv)_{x,y} + (r^*v)_{x,y} = s_{x,y}. \end{aligned} \tag{4.18}$$

The initial values of $(qv)_x$ are *$(\bar{q}\bar{v})_x$, those of $(qv)_y$ are determined from (4.17).*

Theorem 4.4 does not simply follow from Theorem 4.3 because $r = k\hat{q}q_x + \hat{q}q_y + r^*$ need not possess continuous derivatives. Therefore, we first approximate q uniformly in $\mathcal{R}$ by functions $q^{(\sigma)}$ which have continuous second derivatives, such that also the first derivatives of q are approximated by those of $q^{(\sigma)}$, uniformly in $\mathcal{R}$. By virtue of Theorem 4.2 a function $v^{(\sigma)}$ exists in $\mathcal{R}$ which satisfies equation $(4.17)^\sigma$ (in obvious notation) and assumes the initial value $\bar{v}$ on $\mathcal{J}$. Since the coefficients $r^\sigma = k\hat{q}^{(\sigma)}q_x^{(\sigma)} + \hat{q}^{(\sigma)}q_y^{(\sigma)} + r^*$ and s of $(4.17)^\sigma$ possess continuous derivatives, Theorem 4.3 can be applied. It yields that $v^{(\sigma)}$ possesses continuous derivatives $v_{x,y}{}^{(\sigma)}$ which satisfy by (4.14) equation

$$k(v_{x,y}{}^{(\sigma)})_x + v_{x,y}{}^{(\sigma)} + k_{x,y}v_x{}^{(\sigma)} + ((k\hat{q}^{(\sigma)}q_x{}^{(\sigma)} + q^{(\sigma)}q_y{}^{(\sigma)})v^{(\sigma)})_{x,y} + (r^*v)_{x,y} = s_{x,y}.$$

This equation can be written in the form

$$\begin{aligned} k(\hat{q}^{(\sigma)}(q^{(\sigma)}v^{(\sigma)})_{x,y})_x &+ (\hat{q}^{(\sigma)}(q^{(\sigma)}v^{(\sigma)})_{x,y})_y \\ &+ k_{x,y}\hat{q}^{(\sigma)}(q^{(\sigma)}v^{(\sigma)})_x + k\hat{q}_{x,y}{}^{(\sigma)}(q^{(\sigma)}v^{(\sigma)})_x \\ &+ \hat{q}_{x,y}{}^{(\sigma)}(q^{(\sigma)}v_y{}^{(\sigma)})_y - (k\hat{q}_x{}^{(\sigma)} + \hat{q}_y{}^{(\sigma)})(q^{(\sigma)}v^{(\sigma)})_{x,y} \\ &+ r^*\hat{q}(qv)_{x,y} + (r^*\hat{q})_{x,y}qv = s_{x,y} \end{aligned} \tag{4.18$^\sigma$}$$

as is verified by straight differentiation. One observes that second derivatives of $q^{(\sigma)}$ result only from the first term in the last equation. We apply Lemma

4.4 to the equations $(4.18)^\sigma$ and $(4.17)^\sigma$ for the system of functions $\hat{q}^{(\sigma)}(q^{(\sigma)}v^{(\sigma)})_{x,y}$ and $v^{(\sigma)}$. Since the coefficients of these quantities corresponding to $r^{(\sigma)}$ in Lemma 4.4 involve only first derivatives of $q^{(\sigma)}$ and r^*, they converge uniformly to limit functions. Lemma 4.4 is, therefore, applicable. It yields that $\hat{q}^{(\sigma)}(q^{(\sigma)}v^{(\sigma)})_{x,y}$ and $v^{(\sigma)}$ converge uniformly in $\mathfrak{R}$ to continuous limit functions $w_{x,y}$ and $v^{(\infty)}$, satisfying the differential equations $(4.17)^\infty$ and

$$(4.18)^\infty \quad \begin{aligned} k(w_{x,y})_x + (w_{x,y})_y + k_{x,y}w_x + k\hat{q}_{x,y}qw_x \\ + \hat{q}_{x,y}qw_y - (k\hat{q}_x + \hat{q}_y)qw_{x,y} + r^*w_{x,y} + (r^*\hat{q})_{x,y}qv^\infty \\ = s_{x,y}, \end{aligned}$$

and assuming the initial values $\overline{\hat{q}(qv)_{x,y}}$ and $\bar{v}$ on $\mathfrak{I}$. Since $(q^{(\sigma)}v^{(\sigma)})_{x,y}$ are the derivatives of $q^{(\sigma)}v^{(\sigma)}$ it follows that the functions $qw_{x,y}$ are the derivatives of qv^∞. From the uniqueness Theorem 4.1 applied to $(4.17)^\infty$ we have $v^\infty = v$. It is thus shown that qv possesses continuous derivatives $(qv)_{x,y}$ and that the functions $\hat{q}(qv)_{x,y}$ satisfy equation (4.18). Thus Theorem 4.4 is proved.

5. Linear equations in general form. We now proceed to formulate analogues to the lemmas of **4** for the linear differential equation

$$(5.01) \qquad \Lambda \cdot u + cu = au_x + bu_y + cu = f$$

in which the matrices a and b, c and the vector f are continuous functions of x and y in a domain $\mathfrak{R}$, and do not depend on u. In agreement with the stipulations of **1** we require that matrices p and q are given for which (1.3) and (1.4) hold and which, together with a and b, possess continuous derivatives with respect to x and y in $\mathfrak{R}$. The domain $\mathfrak{R}$ is to be a determinacy domain. A continuous initial function $\bar{u}(x)$ on $\mathfrak{I}$ is to be given.

We say that a continuous function $u(x,y)$ in $\mathfrak{R}$ is a solution of equation (5.01) in the wider sense if the function

$$(5.02) \qquad v = q^{-1}u,$$

cf. (3.08), is a solution in the wider sense of the equation

$$(5.03) \qquad kv_x + v_y + rv = s,$$

cf. (4.01), in which

$$(5.04) \quad r = p(aq_x + bq_y + cq) = kq^{-1}q_x + q^{-1}q_y + r^*, \; r^* = pcq, \; s = pf.$$

We note that the diagonal matrix is continuously differentiable and that r and s are continuous in $\mathfrak{R}$. The results of **4** can therefore be translated to statements concerning the equation (5.01).

Identifying the arbitrary matrix q occurring in Lemma 4.2 with the specific matrix q now under consideration we have from Lemma 4.2

LEMMA 5.1. *Suppose the function u in $\mathfrak{R}$ satisfies the equation (5.01) in the wider sense and assumes the initial value $\bar{u}$ on $\mathfrak{I}$. Then u satisfies for every number $\lambda > 0$ the inequality*

$$(5.05) \qquad [\lambda - JR] \, \| ue^{-\lambda y} \| \leq \lambda(1 + L_\lambda) \| \bar{u} \| + \lambda J \| \bar{u} - \bar{\bar{u}} \| + PQ \| fe^{-\lambda y} \|$$

in which J and L_λ are defined by (2.23) and (2.24) while

$$(5.06) \qquad R = \| p(aq_x + bq_y + cq) \| \leq P(AQ_1 + CQ).$$

The latter inequality follows from (2.06), (2.08), (2.10), and (2.16). Theorem 4.1 goes over into the uniqueness

THEOREM 5.1. *There exists in $\mathfrak{R}$ at most one solution u of equation (5.01) in the wider sense assuming on $\mathfrak{I}$ the given initial value $u = \bar{u}$.*

This theorem also follows immediately from Lemma 5.1. An immediate consequence of Lemmas 4.3 and 4.4 are their analogues:

LEMMA 5.2. *Let $c^{(\sigma)}$ and $f^{(\sigma)}$ be a sequence of continuous functions in $\mathfrak{R}$ such that $\| c^{(\sigma)} \|$ remains bounded while $\| f^{(\sigma)} \|$ approaches zero. Let $u^{(\sigma)}$ be a solution in the wider sense of equation*

$$(5.07) \qquad \wedge \cdot u^{(\sigma)} + c^{(\sigma)} u^{(\sigma)} = f^{(\sigma)}$$

assuming on $\mathfrak{I}$ the initial values zero. Then $\| u^{(\sigma)} \| \to 0$ as $\sigma \to \infty$.

LEMMA 5.3. *Let $c^{(\sigma)}$ and $f^{(\sigma)}$ be a sequence of continuous functions in $\mathfrak{R}$ which converge uniformly in $\mathfrak{R}$ to continuous limit functions c and f. Let $u^{(\sigma)}$ be a solution in the wider sense of the equation (5.07) assuming the initial value $u^{(\sigma)} = \bar{u}$ on $\mathfrak{I}$, converging uniformly in $\mathfrak{R}$ to a continuous limit function u. Then u satisfies equation (5.01) in the wider sense. Incidentally, such a sequence $u^{(\sigma)}$ always does converge uniformly in $\mathfrak{R}$ to a continuous limit function.*

From Theorem 4.2 we obtain the existence theorem

THEOREM 5.2. *There exists in $\mathfrak{R}$ a solution of the equation*

$$(5.01) \qquad \wedge \cdot u + cu = f$$

in the wider sense assuming the given initial value $u = \bar{u}$.

Theorem 4.3 will lead to the differentiability theorem

8

THEOREM 5.3. *Suppose that not only the matrices a, b, p and q but also the matrix c and the vector f possess continuous derivatives in $\mathfrak{R}$, and that the initial function $\bar{u}$ possesses a continuous derivative in $\mathfrak{I}$. Then a solution u of equation (5.01) in the wider sense assuming the initial values $\bar{u}$ on $\mathfrak{I}$ possesses continuous derivatives $u_{x,y}$ in $\mathfrak{R}$ and satisfies equation (5.01) in the strict sense. The derivatives $u_{x,y}$ in $\mathfrak{R}$ satisfy in the wider sense the equation*

$$(5.08) \qquad \wedge \cdot u_{x,y} + \wedge_{x,y} \cdot u + c u_{x,y} + c_{x,y} u = f_{x,y}$$

(using the notation defined through (2.15); note that $\wedge_u = 0$ here). The initial value $\overline{u_x}$ of u_x is $\bar{u}_x$, while the initial value $\overline{u_y}$ of u_y is determined from (5.01). (Note that the determinant of $b = p^{-1}q^{-1}$ does not vanish since those of p and q are finite.)

We first see that Theorem 4.4 may be applied since the function r given by (5.04) is of the form assumed in Theorem 4.4 and, by virtue of the assumptions of Theorem 5.3, the functions r^* and s given by (5.04) are continuously differentiable and also $\bar{v} = q^{-1}\bar{u}$ is continuously differentiable. As a result we have that $v = q^{-1}u$ and hence u possesses continuous derivatives in $\mathfrak{R}$ and that equation (4.18) is satisfied in the wider sense. This equation is equivalent with equation

$$(5.09) \qquad p \wedge \cdot u_{x,y} + k_{x,y} q^{-1} u_x + k q^{-1}{}_{x,y} u_x + q^{-1}{}_{x,y} u_y + (pcu)_{x,y} = (pf)_{x,y}$$

by the definition of the wider sense of the operation $\wedge$. In view of $k = paq$, $1 = pbq$ and the relation $au_x + bu_y + cu = f$ we see that this equation is equivalent with (5.08).

Applying Theorem 5.3 to the pair $u_{x,y}$ and equation (5.08) we immediately obtain

THEOREM 5.4. *Suppose that a, b, p, q, c, and f possess continuous second derivatives in $\mathfrak{R}$, and that $\bar{u}$ possesses a continuous second derivative in $\mathfrak{I}$. Let the function u in $\mathfrak{R}$ be a solution of equation (5.01) in the wider sense assuming the initial value $\bar{u}$ on $\mathfrak{I}$. Then u possesses continuous second derivatives $u_{xx,xy,yy}$ in $\mathfrak{R}$. Equation (5.08) is satisfied in the strict sense, and the second derivatives satisfy in the wider sense the equation*

$$(5.10) \qquad \wedge \cdot u_{xx,xy,yy} + 2\wedge_{x,y} \cdot u_{x,y} + \wedge_{xx,xy,yy} \cdot u + (cu)_{xx,xy,yy} = f_{xx,xy,yy}.$$

The initial value $\overline{u_{xx}}$ of u_{xx} is $\bar{u}_{xx}$; the initial values $\overline{u_{xy}}$, $\overline{u_{yy}}$ of u_{xy} and u_{yy} are determined from (5.08).

The combination of Theorems 5. 1, 5. 2, and 5. 3 corresponds to a theorem derived by Perron [3], who did, however, require the existence of continuous second derivatives of $b^{-1}a$, $b^{-1}c$, and $b^{-1}f$. Instead, we needed to require the existence of only first derivatives of these quantities owing to the special effort embodied in Lemma 4. 4. This fact is quite essential for the application to nonlinear equations.

6. Semi-linear differential equations. We call the equation

$$\Lambda \cdot u = au_x + bu_y = g \tag{6.01}$$

semi-linear or one with fixed characteristics if the matrices a and b are defined as functions of x and y in a domain $\mathcal{R}$ while the right member g is a function of x, y, and u in a domain $\mathcal{R}_\Omega$. Of course, we again assume that a, b, p, and q are continuously differentiable in $\mathcal{R}$ and that $\mathcal{R}$ is a determinacy domain. Of the function $g(x, y, u)$ we assume that it has a continuous derivative g_u with respect to u in $\mathcal{R}_\Omega$, cf. (2.11). The initial function $\bar{u}(x)$ on $\mathcal{J}$ is subjected to the condition

$$\| \bar{u} \| < \Omega; \tag{6.02}$$

it is here convenient to exclude the equality.

We say that a continuous function $u(x, y)$ in $\mathcal{R}_\Omega$ obeying inequality

$$\| u \| \leq \Omega \tag{6.03}$$

satisfies equation (6. 01) in the wider sense if it satisfies in the wider sense the linear equation that results when the function u is inserted in $g(x, y, u)$, thus producing a continuous function $g(x, y, u(x, y))$. Lemma 5. 1 now yields immediately

LEMMA 6. 1. *If the function u obeying inequality* (6. 03) *satisfies equation* (6. 01) *in the wider sense, the estimate*

$$[\lambda - JPAQ_1] \| ue^{-\lambda y} \| \leq \lambda(1 + L_\lambda) \| \bar{u} \| + \lambda J \| \bar{u} - \bar{\bar{u}} \| + PQ \| ge^{-\lambda y} \| \tag{6.04}$$

holds for any positive number $\lambda > 0$. We note that the norm $\| ge^{-\lambda y} \|$ here refers to $\mathcal{R}_\Omega$.

From Theorem 5. 1 we shall derive the uniqueness

THEOREM 6. 1. *There is at most one function u obeying inequality* (6. 03) *satisfying equation* (6. 01) *in the wider sense, and assuming the given initial value $\bar{u}$ on $\mathcal{J}$.*

To do so let $u^{(1)}$ and $u^{(2)}$ be two such solutions. We introduce the function.

$$(6.05) \qquad \tilde{g}_u(x,y) = \int_0^1 g_u(x,y,\alpha u^{(2)}(x,y) + (1-\alpha)u^{(1)}(x,y))\,d\alpha,$$

which is defined and continuous in $\mathcal{R}$ since $\alpha u^{(2)} + (1-\alpha)u^{(1)}$ obeys inequality (6.03). Its significance results from the identity

$$(6.06) \qquad g(x,y,u^{(2)}) - g(x,y,u^{(1)}) = \tilde{g}_u(x,y)\,(u^{(2)} - u^{(1)}).$$

It is then clear that the difference $u = u^{(2)} - u^{(1)}$ satisfies the equation

$$\Lambda \cdot u = \tilde{g}_u u,$$

which may be considered a linear equation. Hence Theorem 5.1 yields the statement.

We now shall characterize *a restricted type of domain* for which we shall prove the existence of a solution. First of all we choose a positive number $\theta < 1$. Then we say a domain $\mathcal{R}$ is "restricted" if a number $\lambda > 0$ can be so found that firstly

$$(6.07) \qquad PQG_u < \theta[\lambda - JPAQ_1],$$

the norm $G_u = \| g_u \|$ here referring to the domain $\mathcal{R}_\Omega$, and secondly,

$$(6.08) \qquad \lambda(1 + L_\lambda)\| \bar{u} \| + \lambda J \| \bar{u} - \bar{\bar{u}} \| + PQG < [\lambda - JPAQ_1]\Omega e^{-\lambda Y},$$

Y being the maximum the ordinate y attains in the domain $\mathcal{R}$.

To show that this restriction is not undue we make the

Remark. In the neighborhood of any point x_0 in $\mathcal{J}$ the "center," a restricted domain

$$\mathcal{R}': \; | x - x_0 | + Ky \leq \epsilon, \quad 0 \leq y \leq Y', \quad x, y \text{ in } \mathcal{R}$$

can be found. That such a domain $\mathcal{R}'$ is a determinacy domain was stated in the "Remark" in 3.

To prove the statement we first choose ϵ and a value Y'' such that the quantities $L''_0 = \| q \|'' \, \| q^{-1} - \bar{q}^{-1} \|''$ and $\| \bar{u} - \bar{\bar{u}} \|''$ referring to the corresponding domain $\mathcal{R}''$, see (2.22), are so small that

$$L''_0 \| \bar{u} \|'' + J'' \| \bar{u} - \bar{\bar{u}} \|'' < \Omega - \| \bar{u} \|'',$$

which is possible by virtue of (6.02). Then we choose λ so large that

$$P''Q''G''_u + \theta J''P''A''Q''_1 < \theta\lambda$$

and

$$L''_0 \| \bar{u} \|'' + J'' \| \bar{u} - \bar{\bar{u}} \|'' + \lambda^{-1} P'' Q'' G'' \\ + \lambda^{-1} J'' P'' A'' Q''_1 \Omega < \Omega - \| \bar{u} \|''.$$

Finally we choose a positive number $Y' \leq Y''$ so small that

$$(1 + (1 - e^{-\lambda Y'}) J'' + L''_0) \| \bar{u} \|'' + J'' \| \bar{u} - \bar{\bar{u}} \|'' + \lambda^{-1} P'' Q'' G'' \\ + \lambda^{-1} J'' P'' A'' Q''_1 \, \Omega e^{-\lambda Y'} < \Omega e^{-\lambda Y'}.$$

In view of $\| \ \|' \leq \| \ \|''$ and $L'_\lambda \leq (1 - e^{-\lambda Y'}) J' + L'_0 \leq (1 - e^{-\lambda Y'}) J'' + L''_0$ we see that relations (6.07) hold with reference to the domain $\mathfrak{R}'$. Thus the statement of the Remark is proved.

We now formulate the existence

THEOREM 6.2. *In a restricted domain $\mathfrak{R}$ there exists a solution of equation* (6.01) *in the wider sense obeying inequality* (6.03) *and assuming the given initial value $u = \bar{u}$ on $\mathfrak{J}$.*

The proof proceeds by iterations. We shall construct a sequence of continuous functions $u^{(\sigma)}$ in $\mathfrak{R}$ obeying (6.02) and assuming the initial values $u^{(\sigma)} = \bar{u}$ on $\mathfrak{J}$. We begin with $u^{(0)} = \bar{u}$. Suppose $u^{(\sigma)}$ were determined; then $g^{(\sigma)} = g(x, y, u^{(\sigma)})$ is a continuous function and the linear equation in the wider sense

$$(6.01)^{(\sigma)} \qquad \wedge \cdot u^{(\sigma+1)} = g^{(\sigma)}$$

has by Theorem 5.2 a continuous solution $u^{(\sigma+1)}$ assuming the value $u^{(\sigma+1)} = \bar{u}$ on $\mathfrak{J}$. From Lemma 5.1 we have

$$[\lambda - JPAQ_1] \| u^{(\sigma+1)} e^{-\lambda y} \| \leq \lambda (1 + L_\lambda) \| \bar{u} \| \\ + \lambda J \| \bar{u} - \bar{\bar{u}} \| + PQ \| g^{(\sigma)} e^{-\lambda y} \|;$$

hence by (6.08)

$$\| u^{(\sigma+1)} e^{-\lambda y} \| < \Omega e^{-\lambda Y},$$

whence

$$(6.09) \qquad \| u^{(\sigma+1)} \| \leq \Omega;$$

i.e., $u^{(\sigma+1)}$ satisfies (6.03). Therefore, the iterations can be carried out. We proceed to prove the convergence.

The difference $\Delta u^{(\sigma+1)} = u^{(\sigma+1)} - u^{(\sigma)}$ satisfies in the wider sense the "linear" equation

$$(6.10) \qquad \wedge \cdot \Delta u^{(\sigma+1)} = \tilde{g}_u^{(\sigma)} \Delta u^{(\sigma)} \text{ in } \mathfrak{R}$$

and assumes the initial value zero on $\mathfrak{J}$. Here

$$(6.11)\qquad \tilde{g}_u{}^{(\sigma)}(x,y) = \int_0^1 g_u(x,y,\alpha u^{(\sigma)}(x,y) + (1-\alpha)u^{(\sigma-1)}(x,y))\,d\alpha$$

is a continuous function of x and y with $\| \tilde{g}_u \| \leq \| g_u \| = G_u$, see (6.05) and (6.06). Therefore Lemma 5.1 yields the estimate

$$[\lambda - JPAQ_1] \; \| \Delta u^{(\sigma+1)} e^{-\lambda y} \| \leq PQG_u \; \| \Delta u^{(\sigma)} e^{-\lambda y} \| ;$$

hence by (6.07)

$$\| \Delta u^{(\sigma+1)} e^{-\lambda y} \| \leq \theta \; \| \Delta u^{(\sigma)} e^{-\lambda y} \|,$$

whence

$$\| \Delta u^{(\sigma)} e^{-\lambda y} \| \leq \theta^{\sigma-1} \; \| \Delta u^{(1)} e^{-\lambda y} \|.$$

Consequently, the functions $u^{(\sigma)}$ converge uniformly in $\mathfrak{R}$ to a continuous limit function u. From (6.09) relation (6.03) evidently follows.

Clearly, the functions $g^{(\sigma)} = g(x,y,u^{(\sigma)})$ then also converge uniformly in $\mathfrak{R}$ to $g(x,y,u)$. Now Lemma 5.3 with $c^{(\sigma)} = 0$, $f^{(\sigma)} = g^{(\sigma)}$ can be applied. It follows that the function u satisfies equation (6.01) in the wider sense. Theorem 6.2 is thus proved.

Next we form the differentiability theorem

THEOREM 6.3. *Suppose that not only a, b, p, and q have continuous derivatives in $\mathfrak{R}$, but that in addition $g(x,y,u)$ has continuous derivatives $g_{x,y}$, g_u in $\mathfrak{R}_\Omega$ and that $\bar{u}$ has a continuous derivative $\bar{u}_x$ in $\mathfrak{I}$. Let u be a solution of equation* (6.01) *in the wider sense in $\mathfrak{R}$ obeying the condition*

$$(6.12)\qquad \| u \| < \Omega$$

(*note in comparing with* (6.03) *that the equality is omitted*), *and assuming the value $u = \bar{u}$ on $\mathfrak{I}$. Then the function u possesses continuous derivatives $u_{x,y}$ in $\mathfrak{R}$ and equation* (6.01) *is satisfied in the strict sense. These derivatives satisfy in the wider sense the equations*

$$(6.13)\qquad \Lambda \cdot u_{x,y} + \Lambda_{.x,y} \cdot u - g_u u_{x,y} = g_{x,y}.$$

The initial value $\overline{u_x}$ of u_x is $\bar{u}_x$; the initial value $\overline{u_y}$ of u_y is determined from (6.01).

For the proof of Theorem 6.3 we shall first impose the additional condition that the derivatives $g_{x,y}$ and g_u possess continuous derivatives with respect to u in $\mathfrak{R}_\Omega$. Also we shall confine ourselves to establishing the statement first for a "restricted" domain $\mathfrak{R}$. As a matter of fact, we modify the definition of "restrictedness" somewhat by requiring

$$(6.14)\qquad PQG_u < \theta[\lambda - JP(AQ_1 + A_1Q)]$$

instead of (6.07). Clearly, the Remark about the existence of restricted domains is valid also for this modified restrictedness.

In the restricted domain $\mathcal{R}$ we set up the same iterations that were employed for the proof of the existence Theorem 6.2. The functions $u^{(\sigma+1)}$ satisfy in the wider sense the equation $(6.01)^{(\sigma)}$. If the function $u^{(\sigma)}$ has continuous derivatives in $\mathcal{R}$ then it follows from Theorem 5.3, the assumptions of which are then satisfied, that also $u^{(\sigma+1)}$ possesses continuous derivatives satisfying in the wider sense the equations

$$\Lambda \cdot u_{x,y}{}^{(\sigma+1)} + \Lambda_{x,y} u^{(\sigma+1)} = g_{x,y}{}^{(\sigma)} + g_u{}^{(\sigma)} u_{x,y}{}^{(\sigma)}. \tag{6.15}$$

We shall show that the functions $u_{x,y}{}^{(\sigma)}$ converge to limit functions $u_{x,y}$ which are the derivatives of the function u in question. To this end we first apply Lemma 5.1 to (6.15) with $c = (a_{x,y}, b_{x,y})$ obtaining

$$\begin{aligned}&[\lambda - JP(AQ_1 + A_1 Q)] \, \| u_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| \\ &\quad \leq \lambda(1 + L_\lambda) \| \overline{u_{x,y}} \| + \lambda J \| \bar{u}_{x,y} - \bar{\bar{u}}_{x,y} \| + PQ\{ \| g_{x,y} e^{-\lambda y} \| + \| g_u \| \, \| u^{(\sigma)} e^{-\lambda y} \| .\end{aligned}$$

By virtue of (6.14) we now obtain with an appropriate constant $\Gamma = \Gamma_\lambda$,

$$\| u_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| \leq \theta \, \| u_{x,y}{}^{(\sigma)} e^{-\lambda y} \| + \Gamma,$$

whence by induction, assuming Γ_λ so large that $\| u^{(1)} e^{-\lambda y} \| \leq \Gamma/(1-\theta)$, $\| u_{x,y}{}^{(\sigma)} e^{-\lambda y} \| \leq \Gamma/(1-\theta)$ or

$$U_1{}^{(\sigma)} = \| u_{x,y}{}^{(\sigma)} \| \leq \Gamma e^{\lambda Y}/(1-\theta). \tag{6.16}$$

Next we consider the differences $\Delta u^{(\sigma+1)} = u^{(\sigma+1)} - u^{(\sigma)}$, $\Delta u_{x,y}{}^{(\sigma+1)} = u_{x,y}{}^{(\sigma+1)} - u_{x,y}{}^{(\sigma)}$. They satisfy the equation

$$\begin{aligned}&\Lambda \cdot \Delta u^{(\sigma+1)} = \tilde{g}_u{}^{(\sigma)} \Delta u^{(\sigma)} \\ &\Lambda \cdot \Delta u_{x,y}{}^{(\sigma+1)} + \Lambda_{x,y} \cdot \Delta u^{(\sigma+1)} \\ &\quad = \tilde{g}_{u,x,y}{}^{(\sigma)} \Delta u^{(\sigma)} + \tilde{g}_{uu}{}^{(\sigma)} \Delta u^{(\sigma)} u_{x,y}{}^{(\sigma)} + g_u{}^{(\sigma-1)} \Delta u_{x,y}{}^{(\sigma)}\end{aligned}$$

in the wider sense; hence we have from Lemma 5.1

$$[\lambda - JPAQ_1] \, \| \Delta u^{(\sigma+1)} e^{-\lambda y} \| \leq PQG_u \, \| \Delta u^{(\sigma)} e^{-\lambda y} \|$$

and

$$\begin{aligned}&[\lambda - JP(AQ_1 - A_1 Q)] \, \| \Delta u_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| \\ &\quad \leq PQG_u \, \| \Delta u_{x,y}{}^{(\sigma)} e^{-\lambda y} \| + PQ(G_{u1} + G_{uu} U_1{}^{(\sigma)}) \, \| \Delta u^{(\sigma)} e^{-\lambda y} \| .\end{aligned}$$

By Lemma 5.1 combined with (6.14) and (6.16) we obtain

$$\| \Delta u^{(\sigma+1)} e^{-\lambda y} \| \leq \theta \, \| \Delta u^{(\sigma)} e^{-\lambda y} \|$$

and

$$\| \Delta u_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| \leq \theta \, \| \Delta u_{x,y}{}^{(\sigma)} e^{-\lambda y} \| + Z \, \| \Delta u^{(\sigma)} e^{-\lambda y} \|$$

with an appropriate constant $Z = Z_\lambda$. The first inequality yields

$$\| \Delta u^{(\sigma)} e^{-\lambda y} \| \leq B\theta^{\sigma+1} \text{ with } B = \| \Delta u^{(0)} e^{-\lambda y} \| .$$

The second inequality yields, by induction,

$$\| \Delta u_{x,y}{}^{(\sigma)} e^{-\lambda y} \| \leq B_1\theta^\sigma + ZB\sigma\theta^{\sigma-1} \text{ with } B_1 = \| \Delta u_{x,y}{}^{(1)} e^{-\lambda y} \| .$$

Since $\sum_{\sigma=0}^{\infty} \theta^\sigma < \infty$ and $\sum_{\sigma=0}^{\infty} \sigma\theta^{\sigma-1} < \infty$, it follows that $\| (u^{(\sigma)} - u^{(\tau)}) e^{-\lambda y} \| \to 0$ and $\| (u_{x,y}{}^{(\sigma)} - u_{x,y}{}^{(\tau)}) e^{-\lambda y} \| \to 0$ as $\sigma, \tau \to \infty$. Therefore $u^{(\sigma)}$ and $u_{x,y}{}^{(\sigma)}$ approach uniformly in $\mathcal{R}$ continuous limit functions u^∞ and $u_{x,y}{}^\infty$. Clearly, $u_x{}^\infty$ and $u_y{}^\infty$ are derivatives of u^∞. It follows from Lemma 5.3 that u^∞ and $u_{x,y}{}^\infty$ satisfy equations (6.01) and (6.13) in the wider sense assuming the initial values $\bar{u}$ and $\overline{u_{x,y}}$. Theorem 6.1 gives $u^\infty = u$ and hence Theorem 6.3 is proved under the restrictive conditions first introduced.

The combination of Theorems 6.1, 6.2, and 6.3 corresponds to a statement made by Perron [3]; also the proofs given above are similar to his. Perron requires, however, that g and a possess continuous second derivatives, assuming $b = 1$. For the application to the quasi-linear equations in the next section it is necessary to free oneself from these restrictions. The special efforts made to prove Lemma 4.4 and hence Theorem 5.3 made it unnecessary to require the existence of second derivatives of a or q. We now proceed to free ourselves from the restrictive condition, so far employed in the proof of Theorem 6.3, that the derivatives g_x, g_y should possess continuous derivatives with respect to u.

To this end we approximate the function $g(x, y, u)$ with continuous g_x, g_y, g_u uniformly in $\mathcal{R}_\Omega$ by functions $g^{(\sigma)}(x, y, u)$ possessing continuous derivatives $g_{xu}{}^{(\sigma)}$, $g_{yu}{}^{(\sigma)}$, $g_{uu}{}^{(\sigma)}$ in $\mathcal{R}$. Inequalities (6.14) and (6.08) are then satisfied with $g^{(\sigma)}$ instead of g, if necessary after omitting a finite number of the functions $g^{(\sigma)}$. Consequently, solutions $u^{(\sigma)}$ of $\wedge \cdot u^{(\sigma)} = g^{(\sigma)}(x, y, u^{(\sigma)})$ obeying (6.03) and assuming $u^{(\sigma)} = \bar{u}$ on $\mathcal{J}$ exist in $\mathcal{R}$ by Theorem 6.2. It is immediately seen that the functions $u^{(\sigma)}$ converge uniformly in $\mathcal{R}$ to the solution u of $\wedge \cdot u = g$ under consideration; for, the difference $u^{(\sigma)} - u$ satisfies the equation

$$\wedge \cdot (u^{(\sigma)} - u) - \tilde{g}_u{}^{(\sigma)} (u^{(\sigma)} - u) = g^{(\sigma)}(x, y, u) - g(x, y, u)$$

with the initial value zero and therefore Lemma 5.2 is applicable.

On the other hand Theorem 6.3 as far as proved is applicable to $u^{(\sigma)}$. Consequently, $u^{(\sigma)}$ possesses continuous derivatives $u_{x,y}{}^{(\sigma)}$ which satisfy equations (6.13) with $g^{(\sigma)}$ instead of g. The terms $g_{x,y}{}^{(\sigma)}(x, y, u^{(\sigma)})$ and

$g_u^{(\sigma)}(x, y, u^{(\sigma)})$ entering this equation are known to converge uniformly in $\mathcal{R}$ to $g_{x,y}$ and g_u since $\| u^{(\sigma)} - u \| \to 0$. Therefore, Lemma 5. 3 can be applied. It yields that $u_{x,y}{}^{(\sigma)}$ converges uniformly in $\mathcal{R}$ to a function $u_{x,y}{}^{\infty}$ which satisfies equation (6. 13). Evidently u_x and u_y are the derivatives of u. Thus Theorem 6. 3 is proved without requiring existence of $g_{ux,y}$ and g_{uu}.

We must now free ourselves from the restriction to "restricted" domain $\mathcal{R}$. We consider a given solution u of (5. 01) satisfying the assumptions of Theorem 6. 3. It was shown in the "Remark" earlier in this section that every point x_0 of the initial segment $\mathcal{J}$ can be made the center of a segment $\mathcal{J}'$ which is the initial segment of a restricted domain $\mathcal{R}'$. (It was observed earlier that it makes no difference that we have now defined restrictedness somewhat differently by using (6. 14) instead of (6. 07)). Instead of referring the construction of such domains $\mathcal{R}'$ to the initial segment $\mathcal{J}$ on $y = 0$ we may refer it to the segment cut out of $\mathcal{R}$ by any line $y = \text{const.}$ From the definition of a restricted domain (6. 14) and (6. 08), it is seen that an "altitude" Y and a "width" ϵ can be found, that for every point (x_0, y_0) of $\mathcal{R}$ the domain

$$| x - x_0 | + K(y - y_0) \leq \epsilon, \quad y_0 \leq y \leq y_0 + Y_1(x, y) \text{ in } \mathcal{R}$$

is restricted. It then follows that the region $\mathcal{R}$ can be covered by a finite number of such restricted domains $\mathcal{R}'$ such that every point of $\mathcal{R} - \mathcal{J}$ lies in at least one of the regions $\mathcal{R}' - \mathcal{J}'$. For those domains $\mathcal{R}'$ for which $\mathcal{J}'$ lies on $\mathcal{J}$ Theorem 5. 3 is proved; hence u has continuous derivatives in the strip covered by them. This strip certainly contains the initial segment $\mathcal{J}'$ with the smallest value of y of the "shifted" restricted domains. On this new segment $\mathcal{J}'$, therefore, the new initial value $\bar{u}$ has a continuous derivative $u_{x,y}$. Also $\| u \|' < \Omega$ there (here we make use of the condition (6. 12), excluding the equality sign). Hence Theorem 6. 3 as far as proved can be applied; it follows, in particular, that u has continuous derivatives in the adjacent domain $\mathcal{R}'$. So continuing, the statement of 6. 3 can be established for all of the domain $\mathcal{R}$. Theorem 6. 3 is now completely proved.

7. The general quasi-linear differential equation. We now consider the differential equation

$$\Lambda \cdot u = a u_x + b u_y = g, \tag{7. 01}$$

in which the matrices a and b depend on the unknown function u. We assume that a, b, and g are defined as functions of x, y, u in a region on $\mathcal{R}_\Omega$. We assume that a, b and g possess continuous derivatives with respect to x, y, and

u in $\mathcal{R}_\Omega$; in addition we require that transformations p and q are given for every x, y, u in $\mathcal{R}_\Omega$ which also possess continuous derivatives with respect to x, y, and u. The norms $\| a \|$, $\| b \|$, $\| p \|$, $\| q \|, \cdots$ refer to the domain $\mathcal{R}_\Omega$.

We prescribe an initial function $\bar{u}(x)$ on $\mathfrak{I}$ obeying the inequality

$$\| \bar{u} \| < \Omega. \tag{7.02}$$

Suppose that a function $u = u^*$ is defined in $\mathcal{R}$ with continuous derivatives and obeying the inequality $\| u^* \| \leq \Omega$. Then $a^* = a(x, y, u^*(x, y))$ and $b^* = b(x, y, u^*(x, y))$ are functions of x and y in $\mathcal{R}$ with continuous derivatives; the same is true for p and q. If $\mathcal{R}$ is a domain of determinacy with reference to a^*, b^*, we say that $\mathcal{R}$ is a domain of determinacy for u^*. If $\mathcal{R}$ is a region of determinacy for all such functions u we call it a "common" domain of determinacy.

We now formulate the uniqueness

THEOREM 7.1. *Suppose two functions $u^{(0)}$ and $u^{(1)}$ with continuous derivatives defined in $\mathcal{R}$ and obeying the inequality*

$$\| u \| \leq \Omega, \tag{7.03}$$

satisfy the differential equation (7.01) *in the strict sense and assume the same initial value $u^{(0)} = u^{(1)} = \bar{u}$ on $\mathfrak{I}$. Assume that $\mathcal{R}$ is a dependency domain for the function $u^{(0)}$. Then $u^{(1)} = u^{(0)}$.*

Both functions $u = u^{(0)}$ and $u = u^{(1)}$ evidently satisfy equation

$$\wedge^{(0)} \cdot u = a^{(0)} u_x + b^{(0)} u_y = g - (a - a^{(0)}) u_x^{(1)} - (b - b^{(0)}) u_y^{(1)},$$

in which $a^{(0)} = a(x, y, u^{(0)}(x, y))$, $b^{(0)} = b(x, y, u^{(0)}(x, y))$. The right member is by assumption known to be a continuous function of x, y, and u, and to have a continuous derivative with respect to u. Hence Theorem 6.1 yields the statement.

To illustrate the significance of the formulation of Theorem 7.1 we note that it applies when $\mathcal{R}$ is the "domain of dependence" of a point (x, y) for the solution $u^{(0)}$. It then shows that any other solution of the same problem, however the data may differ outside of $\mathcal{R}$, agrees with $u^{(0)}$ in $\mathcal{R}$ and determines, therefore, the same domain of dependence.

Next we introduce the notion of "restricted" domain suitable for the purposes of this section. In agreement with the stipulations of **2** we employ for any function $f(x, y, u)$ the notations

$$\| f \| = F, \qquad \| f_{x,y} \| = F_1, \qquad \| f_u \| = F_u,$$
$$\| f_{xx,xy,yy} \| = F_2, \qquad \| f_{uxy} \| = F_{u1}, \qquad \| f_{uu} \| = F_{uu},$$

in particular

$$\| a_{x,y} \| + \| b_{x,y} \| = A_1, \qquad \| a_u \| + \| b_u \| = A_u, \text{ and so on.}$$

All norms refer to $\mathcal{R}_\Omega$; in particular this is the case for $J = \| q^{-1} \| \, \| q \|$, and $L_\lambda = \| q \| \, \| q^{-1} - \bar{q}^{-1} e^{-\lambda y} \|$, see (2.23), (2.24).

The initial values u were prescribed as to obey (7.02); for the initial values $\overline{u_x} = \bar{u}_x$ of u_x and $\overline{u_y}$, to be calculated from (7.01), a number Ω_1 is to be so found that

$$\| \overline{u_{x,y}} \| < \Omega_1 \tag{7.04}$$

holds. We choose a positive number $\theta < 1$ and say that a domain $\mathcal{R}$ is "restricted" if a number $\lambda > 0$ can be so found that

$$PQ(G_u + A_u\Omega_1) < \theta[\lambda - JPA(Q_1 + Q_u\Omega_1) - JP(A_1 + A_u\Omega_1)Q] \tag{7.05}$$

$$\lambda(1 + L_\lambda) \| \bar{u} \| + \lambda J \| \bar{u} - \bar{\bar{u}} \| + PQG < [\lambda - JPA(Q_1 + Q_u\Omega_1]\Omega_1 e^{-\lambda y} \tag{7.06}$$

$$\lambda(1 + L_\lambda) \| \overline{u_{x,y}} \| + \lambda J \| \overline{u_{x,y}} - \bar{\bar{u}}_{x,y} \| + PQ(G_1 + G_u\Omega_1) < [\lambda - JPA(Q_1 + Q_u\Omega_1) - JP(A_1 + A_u\Omega_1)Q]\Omega_1 e^{-\lambda Y} \tag{7.07}$$

$$\lambda - JPA(Q_1 + Q_u\Omega_1) - 2JP(A_1 + A_u\Omega_1)Q - PQA_u\Omega_1 - PQG_u > 0. \tag{7.08}$$

We should like to emphasize that a domain is characterized as restricted independently of a bound for the second derivatives. By virtue of condition (7.08) we may now introduce a positive number Ω_2, which will serve as bound for the second derivatives, such that

$$\begin{aligned} &\lambda(1 + L_\lambda) \| \overline{u_{xx,xy,yy}} \| + \lambda J \| \overline{u_{xx,xy,yy}} - \overline{\overline{u_{xx,xy,yy}}} \| \\ &\quad + PQ[A_2 + 2A_{u1}\Omega_1 + A_{uu}\Omega_1^2 + A_u\Omega_2 e^{-\lambda Y}]\Omega_1 \\ &\quad + PQ[G_2 + 2G_{u1}\Omega_1 + G_{uu}\Omega_1^2 + G_u\Omega_2 e^{-\lambda Y}] \\ &\quad = [\lambda - JPA(Q_1 + Q_u\Omega_1) - 2JP(A_1 + A_u\Omega_1)]\Omega_2 e^{-\lambda Y}. \end{aligned} \tag{7.09}$$

Suppose that $\mathcal{R}$ is a common dependency domain which is not restricted. Then we should convince ourselves that in the neighborhood of each point x_0 on $\mathcal{I}$ a restricted domain can be found. Accordingly we make the

Remark. Let $\mathcal{R}$ be a common dependency domain. In the neighborhood of each point x_0 in $\mathcal{I}$ a restricted domain

$$\mathcal{R}': \ | x - x_0 | + Ky \leq \epsilon, \ 0 \leq y \leq Y', \ x, y \text{ in } \mathcal{R}$$

can be found. That such a domain $\mathcal{R}'$ is a common determinacy domain

follows from the "Remark" in **3** and the fact that the norm $K = \| k \|$, referring to $\mathcal{R}_\Omega$, is independent of u.

To prove the statement we first choose the numbers $\epsilon > 0$ and $Y'' > 0$ such that in the corresponding domain $\mathcal{R}''$, the quantities $L''_0 = \| q^{-1} \|'' \| q - \bar{q} \|$, $\| \bar{u} - \bar{\bar{u}} \|''$, and $\| \overline{u_{x,y}} - \overline{\overline{u_{x,y}}} \|$ are so small that

$$L''_0 \| \bar{u} \|'' + J'' \| \bar{u} - \bar{\bar{u}} \|'' < \Omega - \| \bar{u} \|''$$

and

$$L''_0 \| \overline{u_{x,y}} \|'' + J'' \| \overline{u_{x,y}} - \overline{\overline{u_{x,y}}} \| < \Omega_1 - \| \overline{u_{x,y}} \|'',$$

which is possible by (7.02) and (7.04). Then we choose λ so large that the inequalities hold that result when in (7.05) to (7.08) all norms are referred to $\mathcal{R}''$, L_λ is replaced by L''_0, and the factor $e^{-\lambda Y}$ is omitted. Finally a positive number $Y' < Y''$ can be so chosen that these relations hold if the factors $e^{-\lambda Y'}$ instead of $e^{-\lambda Y}$ and $(1 - e^{-\lambda Y'}) J'' + L''_0$ instead of L_λ are inserted; clearly these relations then also hold if the norms refer to the domain $\mathcal{R}'$. This domain therefore is restricted.

We now formulate the main existence theorem of the present paper.

THEOREM 7.2. *Suppose that a, b, p, q, and g have continuous second derivatives with respect to x, y, u in $\mathcal{R}_\Omega$, and that the initial function $\bar{u}$ obeys* (6.02) *and has a continuous second derivative with respect to x on $\mathcal{J}$. Assume that $\mathcal{R}$ is a restricted common determinacy domain. Then there exists a function u in $\mathcal{R}$ obeying inequality* (6.03) *and possessing continuous second derivatives in $\mathcal{R}$, which satisfies in the strict sense the differential equation*

$$\Lambda \cdot u = g \tag{7.01}$$

and assumes the initial values $u = \bar{u}$ on $\mathcal{J}$. The derivatives $u_{x,y}$ satisfy the equation

$$\Lambda \cdot u_{x,y} + \Lambda_{x,y} \cdot u + (\Lambda_u u_{x,y}) \cdot u = g_{x,y} + g_u u_{x,y}. \tag{7.10}$$

The initial value on $\mathcal{J}$ of u_x is $\bar{u}_x$, that of u_y is determined from (7.01). *The second derivatives satisfy in the wider sense the equation*

$$\begin{aligned} \Lambda \cdot u_{xx,xy,yy} &+ 2(\Lambda_{x,y} + \Lambda_u u_{x,y}) \cdot u_{x,y} \\ &+ (\Lambda_{xx,xy,yy} + 2\Lambda_{ux,y} u_{x,y} + \Lambda_{uu} u_{x,y} u_{x,y} + \Lambda_u u_{xx,xy,yy}) \cdot u \\ &= g_{x,y} + 2 g_{ux,y} u_{x,y} + g_{uu} u_{x,y} u_{x,y} + g_u u_{xx,xy,yy}. \end{aligned} \tag{7.11}$$

The initial value of u_{xx} is $\bar{u}_{xx}$ on $\mathcal{J}$, those of u_{xy} and u_{yy} are determined from (7.10).

The proof proceeds by iterations. Beginning with $u^{(0)} = \bar{u}$ we construct a sequence of functions $u = u^{(\sigma)}$ with continuous second derivatives and obeying the inequalities

$$(7.12) \qquad \| ue^{-\lambda y} \| \leq \Omega e^{-\lambda Y}$$

$$(7.13) \qquad \| u_{x,y} e^{-\lambda y} \| \leq \Omega_1 e^{-\lambda Y}$$

$$(7.14) \qquad \| u_{xx,xy,yy} e^{-\lambda y} \| \leq \Omega_2 e^{-\lambda Y}.$$

As a consequence of these inequalities we have

$$(7.03) \qquad \| u \| \leq \Omega$$

$$(7.15) \qquad \| u_{x,y} \| \leq \Omega_1,$$

and

$$(7.16) \qquad \| u_{xx,xy,yy} \| \leq \Omega_2.$$

Suppose that $u^{(\sigma)}$ is determined so as to satisfy all these conditions. By virtue of $\| u^{\sigma} \| \leq \Omega$ we may insert $u^{(\sigma)}$ in the coefficients a, b, and g, which thus become functions of x and y possessing continuous second derivatives. Then we can determine a function $u^{(\sigma+1)}$ as the solution of the equation

$$(7.17) \qquad \wedge^{(\sigma)} \cdot u^{(\sigma+1)} = g^{(\sigma)},$$

in obvious notation, assuming the values $\bar{u}$ on $\mathfrak{J}$. Such a solution exists by virtue of Theorem 5.2. Theorems 5.3 and 5.4 assert that $u^{(\sigma+1)}$ has continuous first and second derivatives satisfying the differential equations

$$(7.18) \qquad \wedge^{(\sigma)} \cdot u_{x,y}^{(\sigma+1)} + \wedge_{x,y}^{(\sigma)} \cdot u^{(\sigma+1)} + (\wedge_u^{(\sigma)} u_{x,y}^{(\sigma)}) \cdot u^{(\sigma+1)} = g_{x,y}^{(\sigma)} + g_u^{(\sigma)} u_{x,y}^{(\sigma)}$$

in the strict sense, and

$$(7.19) \qquad \begin{aligned} &\wedge^{(\sigma)} \cdot u_{xx,xy,yy}^{(\sigma+1)} + 2\wedge_{x,y}^{(\sigma)} \cdot u_{x,y}^{(\sigma+1)} + \wedge_{xx,xy,yy}^{(\sigma)} \cdot u^{(\sigma+1)} \\ &+ 2(\wedge_u^{(\sigma)} u_{x,y}^{(\sigma)}) \cdot u_{x,y}^{(\sigma+1)} + 2(\wedge_{ux,y}^{(\sigma)} u_{x,y}^{(\sigma)}) \cdot u^{(\sigma+1)} \\ &+ (\wedge_{uu}^{(\sigma)} u_{x,y}^{(\sigma)} u_{x,y}^{(\sigma)}) \cdot u^{(\sigma+1)} + (\wedge_u^{(\sigma)} u_{xx,xy,yy}^{(\sigma)}) \cdot u^{(\sigma-1)} \\ &= g_{xx,xy,yy}^{(\sigma)} + 2g_{ux,y}^{(\sigma)} u_{x,y}^{(\sigma)} + g_{uu}^{(\sigma)} u_{x,y}^{(\sigma)} u_{x,y}^{(\sigma)} + g_u^{(\sigma)} u_{xx,xy,yy}^{(\sigma)} \end{aligned}$$

in the wider sense. The initial values of $u_x^{(\sigma+1)}$ and $u_{xx}^{(\sigma+1)}$ are $\bar{u}_x$ and $\bar{u}_{xx}$, while those of $u_y^{(\sigma+1)}$, $u_{xy}^{(\sigma+1)}$, $u_{yy}^{(\sigma+1)}$ are to be determined from (7.17) and (7.18), or (7.01) and (7.10).

First of all we must show that the function $u^{(\sigma+1)}$ obeys (7.12), (7.13), (7.14). Applying Lemma 5.1 to equation (7.17) and making use of the assumed fact that $u^{(\sigma)}$ obeys (7.15), $\| u_{x,y}^{\sigma} \| \leq \Omega_1$, we obtain

$$\begin{aligned} &[\lambda - JPA(Q_1 + Q_u\Omega_1)] \, \| u^{(\sigma+1)} e^{-\lambda y} \| \\ &\qquad \leq \lambda(1 + L_\lambda) \, \| \bar{u} \| + \lambda J \, \| \bar{u} - \bar{\bar{u}} \| + PQG, \end{aligned}$$

whence it follows by (7.06) that $u^{(\sigma+1)}$ obeys (7.12) and hence (7.03).

Note that the boundedness of $\| u_{x,y}{}^{(\sigma)} \|$ is used in order to prove the boundedness of $\| u^{(\sigma+1)} \|$.

Applying Lemma 5.1 to equation (7.18) and making use of the assumed fact that $u^{(\sigma)}$ obeys (7.15), $\| u_{x,y}{}^{(\sigma)} \| \leq \Omega_1$, we obtain

$$[\lambda - JPA(Q_1 + Q_u\Omega_1) - JP(A_1 + A_uU_1)] \, \| u_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| \\ \leq \lambda(1 + L_\lambda) \| \overline{u_{x,y}} \| + \lambda J \| \overline{u_{x,y}} - \overline{\overline{u_{x,y}}} \| + PQ(G_1 + G_u\Omega_1),$$

whence it follows by (7.07) that $u^{(\sigma+1)}$ obeys (7.13) and hence (7.15). Note that the boundedness of $\| u_{xx,xy,yy}{}^{(\sigma)} \|$ is not used here.

Applying Lemma 5.1 to equation (7.19) and making use of the assumed fact that $u^{(\sigma)}$ satisfies (7.15) and (7.14) and the already proved fact that $u^{(\sigma+1)}$ satisfies (7.15), we obtain

$$[\lambda - JPA(Q_1 + Q_u\Omega_1) - 2JP(A_1 + A_u\Omega_1)] \, \| u_{xx,xy,yy}{}^{(\sigma+1)} e^{-\lambda y} \| \\ \leq \lambda(1 + L_\lambda) \| \overline{u_{xx,xy,yy}} \| + \lambda J \| \overline{u_{xx,xy,yy}} - \overline{\overline{u_{xx,xy,yy}}} \| \\ + PQ[A_2 + 2A_{u1}\Omega_1 + A_{uu}\Omega_1{}^2 + A_u\Omega_2 e^{-\lambda Y}]\Omega_1 \\ + PQ[G_2 + 2G_{u1}\Omega_1 + G_{uu}\Omega_1{}^2 + G_u\Omega_2 e^{-\lambda Y}],$$

whence it follows by (7.09) that $u^{(\sigma+1)}$ obeys (7.13) and hence (7.16). Thus it is shown that $u^{(\sigma+1)}$ enjoys the properties required in order that the iterations can be carried out.

Next we prove the convergence of $u^{(\sigma)}$ to a limit function. To this end we consider the differences $\Delta u^{(\sigma+1)} = u^{(\sigma+1)} - u^{(\sigma)}$. They satisfy the equation

$$\wedge^{(\sigma)} \cdot \Delta u^{(\sigma+1)} = - (\tilde{\wedge}_u{}^{(\sigma)} \Delta u^{(\sigma)}) \cdot u^{(\sigma)} \tilde{g}_u{}^{(\sigma)} \Delta u^{(\sigma)},$$

in obvious notation, and assume the initial value zero. Note that the coefficient of $\Delta u^{(\sigma)}$ involves the first derivatives $u_{x,y}{}^{(\sigma)}$. Since $\| u_{x,y}{}^{(\sigma)} \| \leq \Omega_1$ was proved we can apply Lemma 5.1 and obtain

$$[\lambda - JPA(Q_1 + Q_u\Omega_1)] \, \| \Delta u^{(\sigma+1)} e^{-\lambda y} \| \leq PQ[G_u + A_u\Omega_1] \, \| \Delta u^{(\sigma)} e^{-\lambda y} \|,$$

hence by (7.05)

$$\| \Delta u^{(\sigma+1)} e^{-\lambda y} \| \leq \theta \, \| \Delta u^{(\sigma)} e^{-\lambda y} \|,$$

whence with $B = \| \Delta u^{(1)} e^{-\lambda y} \|$

$$\| \Delta u^{(\sigma)} e^{-\lambda y} \| \leq B\theta^{\sigma-1}. \qquad (7.20)$$

It then follows that $u^{(\sigma)}$ converges uniformly in $\mathfrak{R}$ to a limit function u.

Similarly, we deduce from (7.18) that the derivatives $\Delta u_{x,y}{}^{(\sigma+1)}$ satisfy the equation

$$\begin{aligned}\triangle^{(\sigma)} \cdot \Delta u_{x,y}{}^{(\sigma+1)} &+ \wedge_{x,y}{}^{(\sigma)} \cdot \Delta u^{(\sigma+1)} + (\wedge_u{}^{(\sigma)} u_{x,y}{}^{(\sigma)}) \cdot \Delta u^{(\sigma+1)} \\ &= -(\bar{\wedge}_u{}^{(\sigma)} \Delta u^{(\sigma)}) \cdot u_{x,y}{}^{(\sigma)} - (\bar{\wedge}_{ux,y}{}^{(\sigma)} \Delta u^{(\sigma)}) \cdot u^{(\sigma)} \\ &\quad -(\bar{\wedge}_{uu}{}^{(\sigma)} \Delta u^{(\sigma)} u_{x,y}{}^{(\sigma)}) \cdot u^{(\sigma)} - (\wedge_u^{(\sigma-1)} \Delta u_{x,y}{}^{(\sigma)}) \cdot u^{(\sigma)} \\ &\quad + \tilde{g}_{ux,y}{}^{(\sigma)} \Delta u^{(\sigma)} + \tilde{g}_{uu}{}^{(\sigma)} \Delta u^{(\sigma)} u_{x,y}{}^{(\sigma)} + g_u{}^{(\sigma-1)} \Delta u_{x,y}{}^{(\sigma)}.\end{aligned}$$

Note that the coefficient of $\Delta u^{(\sigma)}$ in the first term of the right member involves the second derivatives $u_{xx,xy,yy}{}^{(\sigma)}$. Since $\| u_{xx,xy,yy}{}^{(\sigma)} \| \leq \Omega_2$ was proved we can apply Lemma 5.1 and obtain

$$\begin{aligned}[\lambda - JPA(Q_1 + Q_u\Omega_1) - JP(A_1 + A_u\Omega_1)] &\| \Delta u_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| \\ \leq PQ[A_u\Omega_1 - G_u] &\| \Delta u_{x,y}{}^{(\sigma)} e^{-\lambda y} \| \\ + PQ[A_u\Omega_2 + A_{u1}\Omega_1 + A_{uu}\Omega_1{}^2 + G_{u1} + G_{uu}\Omega_1] &\| \Delta u^{(\sigma)} e^{-\lambda y} \|,\end{aligned}$$

whence by (7.05)

$$\| \Delta u_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| \leq \theta \| \Delta u_{x,y}{}^{(\sigma)} e^{-\lambda y} \| + \mathrm{Z} \| \Delta u^{(\sigma)} e^{-\lambda y} \|$$

with an appropriate constant $\mathrm{Z} = \mathrm{Z}_\lambda$. Employing (7.20) we find by induction

$$\| \Delta u_{x,y}{}^{(\sigma+1)} e^{-\lambda y} \| \leqq B_1\theta^\sigma + \Gamma B\sigma\theta^{\sigma+1}$$

with

$$B = \| \Delta u^{(1)} e^{-\lambda y} \|, \qquad B_1 = \| \Delta u_{x,y}{}^{(1)} e^{-\lambda y} \|.$$

Hence the uniform convergence of $u_{x,y}{}^{(\sigma)}$ in $\mathcal{R}$ is assured, see the proof of Theorem 6.3. The continuous limit functions $u_{x,y}$ are evidently the derivatives of the limit function u of $u^{(\sigma)}$. Certainly, u and $u_{x,y}$ obey the inequalities (7.03) and (7.15).

We cannot prove in an analogous manner that the second derivatives $u_{xx,xy,yy}{}^{(\sigma)}$ converge; for to this end we would need a bound for the third derivatives of u^σ, which we have not established.

Inserting the limit function u into $a(x, y, u)$, $b(x, y, u)$, and $g(x, y, u)$ we obtain functions possessing continuous first derivatives with respect to x and y in $\mathcal{R}$. In the following we shall denote these functions by a, b, and g without qualification.

Suppose that we had proved only the boundedness of $\| u^{(\sigma)} \|$ and $\| u_{x,y}{}^{(\sigma)} \|$ and the convergence of u^σ; then we would know the continuity of the functions a and b, but not their differentiability, which was required from the outset. Therefore we were compelled to prove the boundedness of $\| u_{xx,xy,yy}{}^{(\sigma)} \|$ and, accordingly, to assume continuous second derivatives of the data.

We proceed to prove that the limit function u satisfies the equation

$\wedge \cdot u = g$. As seen from (7.17), (7.18), the functions $u^{(\sigma)}$, $u_{x,y}{}^{(\sigma)}$ satisfy the equations

$$(7.21) \qquad \wedge \cdot u^{(\sigma+1)} = g^{(\sigma)} - (\wedge^{(\sigma)} - \wedge) \cdot u^{(\sigma+1)}$$

$$(7.22) \qquad \begin{aligned} \wedge \cdot u_{x,y}{}^{(\sigma+1)} + \wedge_{x,y}{}^{(\sigma)} \cdot u^{(\sigma+1)} + (\wedge_u{}^{(\sigma)} u_{x,y}{}^{(\sigma)}) \cdot u^{(\sigma+1)} \\ = g_{x,y}{}^{(\sigma)} + g_u{}^{(\sigma)} u_{x,y}{}^{(\sigma)} - (\wedge^{(\sigma)} - \wedge) \cdot u_{x,y}{}^{(\sigma+1)}. \end{aligned}$$

Since $\| u_{x,y}{}^{(\sigma+1)} \|$ and $\| u_{xx,xy,yy} \|$ were proved to remain bounded we have

$$\| (\wedge^{(\sigma)} - \wedge) \cdot u^{(\sigma+1)} \| \to 0$$

and

$$\| (\wedge^{(\sigma)} - \wedge) \cdot u_{x,y}{}^{(\sigma+1)} \| \to 0 \text{ as } \sigma \to \infty.$$

The two equations (7.21) and (7.22) can therefore be considered linear equations for $u^{(\sigma+1)}$ and $u_{x,y}{}^{(\sigma+1)}$ whose coefficients converge uniformly to limit functions. Therefore, Lemma 5.3 can be applied. It yields that u and $u_{x,y}$ satisfy in the wider sense equations (7.01) $\wedge \cdot u = g$ and (7.10) $\wedge \cdot u_{x,y} + \wedge_{x,y} u + (\wedge_u u_{x,y}) u = g_{x,y} + g_u u_{x,y}$. Since u possesses continuous derivatives, the first equation is even satisfied in the strict sense.

In order to establish that u possesses continuous second derivatives we shall *not* try to prove that the second derivatives of $u^{(\sigma)}$ converge. Instead we apply Theorem 6.3 to equation (7.10) which we consider as a semi-linear equation for $u_{x,y}$, assuming u as a given function with continuous derivatives, and taking $g_{x,y} + g_u u_{x,y} - (\wedge_u u_{x,y}) \cdot u$ as right member. Note that the last term here is quadratic in $u_{x,y}$. Its derivatives with respect to x, y and $u_{x,y}$ are continuous. We also know that a, b, p, q possess continuous first derivatives with respect to x and y since u has already been proved to possess such derivatives. Hence Theorem 6.3 can be applied. It yields that $u_{x,y}$ possesses continuous first derivatives satisfying equation (7.11). Thus Theorem 7.3 is completely proved.

In conclusion we mention that we could easily add a differentiation theorem stating that u possesses continuous third derivatives if the initial values and a, b, g do. Such a theorem is an immediate consequence of Theorem 6.3 or even of 5.3 since the second derivatives enter equation (7.11) only linearly. We refrain from carrying out the details.

NEW YORK UNIVERSITY.

BIBLIOGRAPHY.

[1] H. Lewy, "Ueber das Anfangswertproblem bei einer hyperbolischen nichtlinearen partiellen Differentialgleichung zweiter Ordnung mit zwei unabhängigen Veränderlichen," *Mathematische Annalen*, vol. 98 (1927), pp. 179-191.

[2] K. Friedrichs and H. Lewy, "Das Anfangswertproblem einer beliebigen nichtlinearen hyperbolischen Differentialgleichung beliebiger Ordnung in zwei Variablen. Existenz, Eindeutigkeit und Abhängigkeitsbereich der Lösung," *Mathematische Annalen*, vol. 99 (1928), pp. 200-221.

[3] O. Perron, "Ueber Existenz und Nichtexistenz von Integralen partieller Differentialgleichungssysteme im reellen Gebiet," *Mathematische Zeitschrift*, vol. 27 (1928), pp. 549-564.

[4] A. Haar, "Sur l'unicité des solutions des équations aux dérivées partielles," *Comptes rendus des séances de l'Académie des Sciences*, vol. 187 (1928), p. 23.

[5] A. Haar, "Zur Charakteristikentheorie," *Acta Litterarum Ac Scientiarum*, vol. 4 (1928), pp. 103-114.

[6] A. Haar, "Ueber Eindeutigkeit und Analyzität der Lösungen partieller Differentialgleichungen," *Atti del Congresso Internazionale dei Matematici*, Bologna, vol. 3 (1928), pp. 5-10.

[7] K. Friedrichs and H. Lewy, "Ueber fortsetzbare Anfangsbedingungen bei hyperbolischen Differentialgleichungen in drei Veränderlichen," *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, No. 26 (1932), pp. 135-143.

[8] J. Schauder, "Das Anfangswertproblem einer quasilinearen hyperbolischen Differentialgleichung zweiter Ordnung in beliebiger Anzahl von unabhängigen Veränderlichen," *Fundamenta Mathematicae*, vol. 24 (1935), pp. 213-246.

[9] F. Frankl, "Ueber das Anfangswertproblem für lineare und nichtlineare hyperbolische partielle Differentialgleichungen zweiter Ordnung," *Recueil Mathématique*, vol. 2 (44) (1937), pp. 793-812.

[10] I. Petrowsky, "Ueber das Cauchysche Problem für Systeme von partiellen Differentialgleichungen," *Recueil Mathématique*, vol. 2 (44) (1937), pp. 815-868.

[11] S. Christianovitch, "Le problème de Cauchy pour les équations non linéaires hyperboliques," *Recueil Mathématique*, vol. 2 (44) (1937), pp. 871-897.

[12] J. Schauder, "Cauchysches Problem für partielle Differentialgleichungen erster Ordnung. Anwendung einiger sich auf die Absolutbeträge der Lösungen beziehenden Abschätzungen," *Commentarii Mathematici Helvetici*, vol. 9 (1937), pp. 263-283.

[13] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, vol. 2 (1937).

[14] R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves*. To appear 1948.

[15] M. Cinquini-Cibrario, "Un teorema di esistenza e di unicità per un sistema di equazioni alle derivate parziali," *Annali di Matematica* (4), vol. 24 (1945), pp. 157-175.

GENERALIZED FREE PRODUCTS WITH AMALGAMATED SUBGROUPS.* [1]

By Hanna Neumann.

PART I. Definitions and General Properties.

Introduction. In his paper "Die Untergruppen der freien Produkte" [2] [2] A. Kurosch determines completely the structure of the subgroups of a free product of groups. He also suggests the problem of determining similarly the subgroups of a free product of groups with an amalgamated subgroup.[3] This problem will be solved completely. However, to describe the structure of these subgroups adequately, a generalization of the free product with one amalgamated subgroup is necessary. It is with this generalization that the first part of our investigations is mainly concerned; the problem of determining the subgroups of free products with one amalgamated subgroup and of generalized free products with amalgamated subgroups will be dealt with in the second part.

Denote by $\mathfrak{G}_\alpha$ given groups, finite or infinite in number, each of which contains a subgroup $\mathfrak{U}_\alpha$ isomorphic to a fixed group $\mathfrak{U}$. In the free product of the groups $\mathfrak{G}_\alpha$ we introduce all the relations identifying every pair of elements of groups $\mathfrak{U}_\alpha$ and $\mathfrak{U}_\beta$ which under some fixed isomorphism correspond to the same element of $\mathfrak{U}$. The result is a group $\mathfrak{G}$ which is generated by groups isomorphic to the given groups $\mathfrak{G}_\alpha$; and any two of these groups have in $\mathfrak{G}$ the same meet isomorphic to $\mathfrak{U}$. $\mathfrak{G}$ is called the free product of the groups $\mathfrak{G}_\alpha$ with the amalgamated subgroup $\mathfrak{U}$. It can be loosely described as the largest group generated by the groups $\mathfrak{G}_\alpha$ such that any two of these have the same meet $\mathfrak{U}$.

Now let the given groups $\mathfrak{G}_\alpha$ be such that any two of them, $\mathfrak{G}_\alpha$ and $\mathfrak{G}_\beta$, contain isomorphic subgroups $\mathfrak{U}_{\alpha\beta}$ and $\mathfrak{U}_{\beta\alpha}$, respectively. We form correspondingly the group obtained from the free product of the groups $\mathfrak{G}_\alpha$ by identifying in it all pairs of corresponding elements $u_{\alpha\beta}$ and $u_{\beta\alpha}$ of the groups $\mathfrak{U}_{\alpha\beta}$ and $\mathfrak{U}_{\beta\alpha}$ (under some fixed isomorphism between $\mathfrak{U}_{\alpha\beta}$ and $\mathfrak{U}_{\beta\alpha}$) for every pair α, β of suffixes ($\alpha \neq \beta$). If this group contains subgroups isomorphic

* Received February 13, 1947.

[1] The greater part of the material in this paper was presented as a D. Phil. thesis at the University of Oxford.

[2] Numbers in brackets refer to the list of references at the end of this paper.

[3] Cf. [6]; and [5], p. 41.

to the $\mathfrak{G}_\alpha$ such that any two of them have the meet $\mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}$, we call this group the free product of the groups $\mathfrak{G}_\alpha$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$ (**1**). Certain sufficient conditions for the existence of this generalized free product can be established immediately (**2**). It is also easily seen that the subgroups $\mathfrak{U}_{\alpha\beta}$ must satisfy certain necessary conditions, if the generalized free product of the groups $\mathfrak{G}_\alpha$ is to exist (**3**). But while the free product with one amalgamated subgroup can always be formed if only the factors all contain a subgroup isomorphic to the same group $\mathfrak{U}$, an example in **3** shows that the construction of the generalized free product with amalgamated subgroups is not always possible, even if the obvious necessary conditions are satisfied.

The concepts introduced so far may be looked upon as a special case of a rather more general problem. In **4**[4] we introduce the notion of an "incomplete group."[5] Two questions arise: whether an incomplete group is imbeddable into a group; and if it is imbeddable, whether there exists a "largest" group which contains the incomplete group and is generated by it. This second question will be answered in the affirmative. That the answer to the first question is not always positive, is known.[6] Since the system of groups $\mathfrak{G}_\alpha$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$ must form an incomplete group, if the generalized free product is to exist, the example in **3** provides an instance of a different type of incomplete group which is not imbeddable.

In **5** we derive a necessary and sufficient criterion for the existence of the generalized free product with amalgamated subgroups reducing the general case to a more special (though by no means easier) case. In **6** we describe the structure of the generalized free product in a somewhat different way. This leads to a number of special results and examples (**7**, **8**). These enable us to derive some simple and useful sufficient criteria for the existence of the generalized free product in the case of three factors. However, they also seem to indicate that more general results in this direction can hardly be hoped for (**9**). We conclude the paper with a detailed discussion of the free product of three infinite cycles with amalgamated subcycles (**10**).

1. We begin with the definition of the generalized free product with amalgamated subgroups.

Let α vary over a finite or infinite set A of suffixes; let $\mathfrak{G}_\alpha$ be given groups with the following property: For every suffix $\beta \neq \alpha$ out of A, $\mathfrak{G}_\alpha$ contains a subgroup $\mathfrak{U}_{\alpha\beta}$ which is isomorphic to the subgroup $\mathfrak{U}_{\beta\alpha}$ of $\mathfrak{G}_\beta$.

---

[4] The contents of this paragraph arose out of a suggestion by R. Baer.

[5] The "incomplete group" defined here is similar to H. Brandt's "Gruppoid." Cf. [1].

[6] A. Malcev has constructed an example of a non-imbeddable semi-group. Cf. [4].

Let $\mathfrak{J}_{\alpha\beta}$ be a fixed isomorphism between $\mathfrak{U}_{\alpha\beta}$ and $\mathfrak{U}_{\beta\alpha}$,

$$\mathfrak{J}_{\alpha\beta}(\mathfrak{U}_{\alpha\beta}) = \mathfrak{U}_{\beta\alpha},$$

for all pairs of suffixes α, β of A, such that $\mathfrak{J}_{\beta\alpha} = \mathfrak{J}_{\alpha\beta}^{-1}$.

We denote by $\mathfrak{U}_\alpha$ the subgroup of $\mathfrak{G}_\alpha$ which is generated by all the groups $\mathfrak{U}_{\alpha\beta}$ (α fixed).

For every group $\mathfrak{G}_\alpha$ we choose a system of generators consisting of all the elements $u_{\alpha\beta}$ of all subgroups $\mathfrak{U}_{\alpha\beta}$ of $\mathfrak{G}_\alpha$, supplemented by elements g_α of $\mathfrak{G}_\alpha - \mathfrak{U}_\alpha$ as necessary. Let $\mathfrak{R}_\alpha$ be a system of defining relations for these generators.

Now let $\mathfrak{G}$ be the group generated by all the elements $u_{\alpha\beta}$ and g_α for all pairs of suffixes α, β of A, with a system of defining relations comprising the systems $\mathfrak{R}_\alpha$ for all suffixes α, and besides the system $\mathfrak{J}$ of "identifying relations,"

$$\mathfrak{J}:\quad u_{\alpha\beta} = \mathfrak{J}_{\alpha\beta}(u_{\alpha\beta})$$

for all pairs α, β and all the elements $u_{\alpha\beta}$ of $\mathfrak{U}_{\alpha\beta}$.

The group $\mathfrak{F}$ with the same generators as $\mathfrak{G}$, but with the relations $\mathfrak{R}_\alpha$ for all α of A only, is the free product of the groups $\mathfrak{G}_\alpha$. If we denote the smallest self-conjugate subgroup of $\mathfrak{J}$ containing all the elements $u_{\alpha\beta}u_{\beta\alpha}^{-1}$ (where $u_{\beta\alpha} = \mathfrak{J}_{\alpha\beta}(u_{\alpha\beta})$) by $\mathfrak{N}_{\mathfrak{J}}$, then $\mathfrak{G}$ is isomorphic to the factor-group of $\mathfrak{F}$ with respect to $\mathfrak{N}_{\mathfrak{J}}$,

$$\mathfrak{G} \cong \mathfrak{F}/\mathfrak{N}_{\mathfrak{J}}.$$

If now it is true that

(i) for every α_0 the meet of $\mathfrak{G}_{\alpha_0}$ and $\mathfrak{N}_{\mathfrak{J}}$ in $\mathfrak{F}$ consists of the unit element only (i. e., every relation which follows from the relations of all the systems $\mathfrak{R}_\alpha$ and $\mathfrak{J}$ together, and which involves generators of one group $\mathfrak{G}_{\alpha_0}$ only, follows from the relations of the system $\mathfrak{R}_{\alpha_0}$ only);

(ii) if G_α and G_β are elements of $\mathfrak{G}_\alpha$ and $\mathfrak{G}_\beta$ respectively ($\alpha \neq \beta$), and if the product $G_\alpha G_\beta$ lies in $\mathfrak{N}_{\mathfrak{J}}$, then $G_\alpha = u_{\alpha\beta}$ and $G_\beta = u_{\beta\alpha}$ (i. e., no relation $G_\alpha = G_\beta^{-1}$ other than the relations of $\mathfrak{J}$ follow from all the relations of the systems $\mathfrak{R}_\alpha$ and $\mathfrak{J}$ taken together),

then we call $\mathfrak{G}$ the free product of the groups $\mathfrak{G}_\alpha$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$. We write [7]

$$\mathfrak{G} = \{\overset{*}{\prod_\alpha} \mathfrak{G}_\alpha ; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}.$$

---

[7] In generalization of the current notation $\mathfrak{F} = \overset{*}{\prod_\alpha} \mathfrak{G}_\alpha$ for the ordinary free product; cf. e. g. Kurosch [2].

It should be noted, however, that $\mathfrak{G}$ depends not only on the groups $\mathfrak{G}_\alpha$ and their subgroups $\mathfrak{U}_{\alpha\beta}$, but also on the choice of the isomorphisms $\mathfrak{J}_{\alpha\beta}$.

If all the groups $\mathfrak{U}_{\alpha\beta}$ are isomorphic to one and the same group $\mathfrak{U}$, the conditions (i) and (ii) are always satisfied, as O. Schreier has shown.[8] $\mathfrak{G}$ is then the free product with one amalgamated subgroup,

$$\mathfrak{G} = \{ \prod_{\alpha}^{*} \mathfrak{G}_\alpha ; \mathfrak{U} = \mathfrak{U} \}.^{9}$$

All results derived for the generalized free product with amalgamated subgroups hold therefore in particular for the free product with one amalgamated subgroup (but may, of course, become trivial).

2. The condition (i) of **1** expresses the fact that if we introduce the relations $\mathfrak{J}$ into the free product $\mathfrak{F} = \prod_{\alpha}^{*} \mathfrak{G}_\alpha$, the subgroups $\mathfrak{G}_\alpha$ of $\mathfrak{F}$ do not "collapse," i. e. they are isomorphically represented in $\mathfrak{G}$. We denote these isomorphic images of the groups $\mathfrak{G}_\alpha$ in $\mathfrak{G}$ by the same letter $\mathfrak{G}_\alpha$.

The condition (ii) makes sure that any two of the groups, $\mathfrak{G}_\alpha$ and $\mathfrak{G}_\beta$, have the exact meet $\mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}$ in $\mathfrak{G}$.

Both conditions are, therefore, satisfied, if the groups $\mathfrak{G}_\alpha$ are given as subgroups of a group $\bar{\mathfrak{G}}$ such that in $\bar{\mathfrak{G}}$, $\mathfrak{G}_\alpha \cap \mathfrak{G}_\beta = \mathfrak{U}_{\alpha\beta}$. Hence:

2. 0 THEOREM. *If all groups $\mathfrak{G}_\alpha$ are subgroups of a group $\bar{\mathfrak{G}}$ such that in $\bar{\mathfrak{G}}$: $\mathfrak{G}_\alpha \cap \mathfrak{G}_\beta = \mathfrak{U}_{\alpha\beta}$, and if every relation between elements of all the groups $\mathfrak{G}_\alpha$ follows from the defining relations of the single groups $\mathfrak{G}_\alpha$ and from those relations which express that an element belongs to the meet $\mathfrak{U}_{\alpha\beta}$ of $\mathfrak{G}_\alpha$ and $\mathfrak{G}_\beta$, then the groups $\mathfrak{G}_\alpha$ generate in $\bar{\mathfrak{G}}$ the generalized free product*

$$\mathfrak{G} = \{ \prod_{\alpha}^{*} \mathfrak{G}_\alpha ; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha} \}.$$

This follows immediately from the definition of the generalized free product.

If again the groups $\mathfrak{G}_\alpha$ are subgroups of a group $\mathfrak{G}^*$ such that $\mathfrak{G}_\alpha \cap \mathfrak{G}_\beta = \mathfrak{U}_{\alpha\beta}$ in $\mathfrak{G}^*$, and if the groups $\mathfrak{G}_\alpha$ generate $\mathfrak{G}^*$, then the elements $u_{\alpha\beta}$ and g_α for all indices α, β form a system of generators for $\mathfrak{G}^*$, and the relations $\mathfrak{R}_\alpha$ for all α, and $\mathfrak{J}$, certainly hold in $\mathfrak{G}^*$. $\mathfrak{G}^*$ may, however, need further relations between these generators for its definition. It follows that $\mathfrak{G}^*$ is isomorphic with the factor group of the free product $\mathfrak{F} = \prod_{\alpha}^{*} \mathfrak{G}_\alpha$ with respect to a self-conjugate subgroup $\mathfrak{N}^*$ of $\mathfrak{F}$ which contains $\mathfrak{N}_{\mathfrak{J}}$. Hence

[8] Cf. [6].

[9] For the free product of two groups with one amalgamated subgroup we also write $\mathfrak{G} = \mathfrak{G}_1 \underset{\mathfrak{U}}{*} \mathfrak{G}_2$.

$$\mathfrak{G}^* \cong \mathfrak{F}/\mathfrak{N}^* \cong \mathfrak{F}/\mathfrak{N}_{\mathfrak{Z}}/\mathfrak{N},$$

where

$$\mathfrak{N} = \mathfrak{N}^*/\mathfrak{N}_{\mathfrak{Z}}.$$

By hypothesis, $\mathfrak{N}^*$ satisfies (i) and (ii); so does, *a fortiori*, $\mathfrak{N}_{\mathfrak{Z}}$; i.e., $\mathfrak{G} = \{\overset{*}{\prod_{\alpha}} \mathfrak{G}_\alpha; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$ exists, and $\mathfrak{G}^* \cong \mathfrak{G}/\mathfrak{N}$.

This self-conjugate subgroup $\mathfrak{N}$ of $\mathfrak{G}$ has the following properties:

1. $\mathfrak{G}_\alpha \cap \mathfrak{N} = 1$ for all α;

2. $\mathfrak{N}$ does not contain any elements of the form $G_\alpha G_\beta$, where $G_\alpha \neq 1$ and $G_\beta \neq 1$ $(\alpha \neq \beta)$ are elements of $\mathfrak{G}_\alpha$ and $\mathfrak{G}_\beta$, respectively. Conversely if a self-conjugate subgroup $\mathfrak{N}$ of $\mathfrak{G}$ has these properties, then the factor group $\mathfrak{G}/\mathfrak{N}$ is generated by groups isomorphic with the groups $\mathfrak{G}_\alpha$, such that any two of them have a meet isomorphic with $\mathfrak{U}_{\alpha\beta}$ in $\mathfrak{G}/\mathfrak{N}$.

We call a self-conjugate subgroup $\mathfrak{N}$ of $\mathfrak{G}$ with the properties 1. and 2. "tidy with respect to the factors $\mathfrak{G}_\alpha$," or just "tidy," if no ambiguity is possible.

Then we have proved:

2.1 THEOREM. *If $\mathfrak{G}^*$ is a group which is generated by a finite or infinite number of subgroups $\mathfrak{G}_\alpha$, and if $\mathfrak{G}_\alpha \cap \mathfrak{G}_\beta = \mathfrak{U}_{\alpha\beta}$ in $\mathfrak{G}^*$, then $\mathfrak{G} = \{\overset{*}{\prod_{\alpha}} \mathfrak{G}_\alpha; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$ exists, and $\mathfrak{G}^* \cong \mathfrak{G}/\mathfrak{N}$, where $\mathfrak{N}$ is a tidy self-conjugate subgroup of $\mathfrak{G}$.*[10]

Theorem 2.1 can obviously be used to define the free product of groups with amalgamated subgroups, *viz.* as the largest group $\mathfrak{G}$ which is generated by subgroups isomorphic with the groups $\mathfrak{G}_\alpha$, such that the meet of any two of these subgroups, $\mathfrak{G}_\alpha$ and $\mathfrak{G}_\beta$, is isomorphic with $\mathfrak{U}_{\alpha\beta}$. Here $\mathfrak{G}$ is called "largest" in the sense, that any other group with these properties is a homomorphic image of $\mathfrak{G}$.

3. If the groups $\mathfrak{G}_\alpha$ are not given as subgroups of one and the same larger group, but are any abstract groups, about which it is known only that they contain subgroups $\mathfrak{U}_{\alpha\beta}$ (in $\mathfrak{G}_\alpha$), such that $\mathfrak{U}_{\alpha\beta}$ and $\mathfrak{U}_{\beta\alpha}$ are isomorphic for every pair of suffixes α, β, under what conditions can the free product of the groups $\mathfrak{G}_\alpha$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$ exist?

If $\mathfrak{G} = \{\overset{*}{\prod_{\alpha}} \mathfrak{G}_\alpha; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$ exists, then the meet of any three different factors $\mathfrak{G}_\alpha$, $\mathfrak{G}_\beta$, $\mathfrak{G}_\gamma$ in $\mathfrak{G}$ can be formed in any one of these factors as the meet

[10] Theorem 2.0 is, of course, only a special case of this Theorem 2.1, but it will be used in just that form in Part II.

$$\mathfrak{U}_{\alpha\beta\gamma} = \mathfrak{U}_{\alpha\beta} \cap \mathfrak{U}_{\alpha\gamma} \text{ in } \mathfrak{G}_\alpha,$$

or

$$\mathfrak{U}_{\beta\gamma\alpha} = \mathfrak{U}_{\beta\gamma} \cap \mathfrak{U}_{\beta\alpha} \text{ in } \mathfrak{G}_\beta,$$

or

$$\mathfrak{U}_{\gamma\alpha\beta} = \mathfrak{U}_{\gamma\alpha} \cap \mathfrak{U}_{\gamma\beta} \text{ in } \mathfrak{G}_\gamma.$$

In $\mathfrak{G}$ the three groups $\mathfrak{U}_{\alpha\beta\gamma}$, $\mathfrak{U}_{\beta\gamma\alpha}$, and $\mathfrak{U}_{\gamma\alpha\beta}$ are, therefore, identical. This means that in the original abstract groups $\mathfrak{G}_\alpha$ the subgroups $\mathfrak{U}_{\alpha\beta}$ must be such that also the three-suffix meets $\mathfrak{U}_{\alpha\beta\gamma}$, $\mathfrak{U}_{\beta\gamma\alpha}$, $\mathfrak{U}_{\gamma\alpha\beta}$ are isomorphic; and the isomorphism $\mathfrak{J}_{\alpha\beta}$ between $\mathfrak{U}_{\alpha\beta}$ and $\mathfrak{U}_{\beta\alpha}$ must map $\mathfrak{U}_{\alpha\beta\gamma}$ onto $\mathfrak{U}_{\beta\gamma\alpha}$, and correspondingly for $\mathfrak{J}_{\beta\gamma}$ and $\mathfrak{J}_{\alpha\gamma}$. These isomorphisms must, therefore, be transitive with respect to the three-suffix meets; i. e., the mapping of $\mathfrak{U}_{\alpha\beta\gamma}$ onto $\mathfrak{U}_{\gamma\alpha\beta}$ effected by first applying $\mathfrak{J}_{\alpha\beta}$, then $\mathfrak{J}_{\beta\gamma}$ must be identical with the mapping provided by $\mathfrak{J}_{\alpha\gamma}$. Hence:

3.0. *In order that the generalized free product of the groups $\mathfrak{G}_\alpha$ can be formed, it is necessary that for any three different suffixes α, β, γ the meets $\mathfrak{U}_{\alpha\beta\gamma} = \mathfrak{U}_{\alpha\beta} \cap \mathfrak{U}_{\beta\gamma}$, $\mathfrak{U}_{\beta\gamma\alpha} = \mathfrak{U}_{\beta\gamma} \cap \mathfrak{U}_{\beta\alpha}$, and $\mathfrak{U}_{\gamma\alpha\beta} = \mathfrak{U}_{\gamma\alpha} \cap \mathfrak{U}_{\gamma\beta}$ are isomorphic, and that the isomorphisms $\mathfrak{J}_{\alpha\beta}$, $\mathfrak{J}_{\beta\gamma}$ and $\mathfrak{J}_{\alpha\gamma}$ provide a mapping between them which, moreover is such that for any element $u_{\alpha\beta\gamma}$ of $\mathfrak{U}_{\alpha\beta\gamma}$:*

$$\mathfrak{J}_{\beta\gamma}\{\mathfrak{J}_{\alpha\beta}(u_{\alpha\beta\gamma})\} = \mathfrak{J}_{\alpha\gamma}(u_{\alpha\beta\gamma}).$$

If these conditions are satisfied, then the corresponding conditions for the meets of more than three factors, $\mathfrak{G}_{\alpha'}, \mathfrak{G}_{\alpha''}, \cdots, \mathfrak{G}_{\alpha^{(\nu)}}, \cdots$ say, are automatically satisfied. This follows immediately from the fact that any such meet can be formed in any one of these factors as the meet of all those three-suffix meets whose three suffixes occur amongst the suffixes $\alpha', \alpha'', \cdots, \alpha^{(\nu)}, \cdots$.

The conditions 3.0 really do no more than describe the kind of groups and isomorphisms for which the definition of the generalized free product makes sense. Groups, and isomorphisms, which do not satisfy 3.0, need not concern us at all.

A necessary condition of a different kind follows from the results of the preceding paragraph.

If $\mathfrak{G} = \{\overset{*}{\prod_\alpha} \mathfrak{G}_\alpha; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$ exists, then $\mathfrak{G}$ contains for every α the subgroup $\mathfrak{U}_\alpha$ of $\mathfrak{G}_\alpha$ which is generated by all the groups $\mathfrak{U}_{\alpha\beta}$ with fixed α. Any two of these, $\mathfrak{U}_\alpha$ and $\mathfrak{U}_\beta$, also have meet $\mathfrak{U}_\alpha \cap \mathfrak{U}_\beta = \mathfrak{U}_{\alpha\beta}$ in $\mathfrak{G}$. Hence, by 2.1 the free product $\mathfrak{U} = \{\overset{*}{\prod_\alpha} \mathfrak{U}_\alpha; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$ of the groups $\mathfrak{U}_\alpha$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$ exists. Therefore:

3. 1. *In order that* $\mathfrak{G} = \{ \overset{*}{\prod_a} \mathfrak{G}_a ; \mathfrak{U}_{a\beta} = \mathfrak{U}_{\beta a} \}$ *may exist, it is necessary that* $\mathfrak{U} = \{ \overset{*}{\prod_a} \mathfrak{U}_a ; \mathfrak{U}_{a\beta} = \mathfrak{U}_{\beta a} \}$ *exists.*

We shall see later that this condition is also sufficient (**5**).

In this paragraph we want to show that in the case of the generalized free product it is not sufficient that the groups $\mathfrak{G}_a$ are formally suited for its construction; the conditions 3. 0 may be satisfied, but the generalized free product need not exist.

This is shown by the following example of three groups $\mathfrak{U}_i$ each of which is generated by its subgroups $\mathfrak{U}_{ij}$ and $\mathfrak{U}_{ik}$.

3. 2. *Example.* Let $\mathfrak{U}_1$ be the direct product of two infinite cycles $\{a_1\}$ and $\{b_1\}$; i. e., $\mathfrak{U}_1$ is the group generated by a_1 and b_1 with the defining relation $a_1b_1 = b_1a_1$. If we put $\mathfrak{U}_{12} = \{a_1\}$ and $\mathfrak{U}_{13} = \{b_1\}$, then $\mathfrak{U}_{12} \cap \mathfrak{U}_{13} = 1$.

Let $\mathfrak{U}_2$ be the group generated by a_2 and c_2 with the defining relation $c_2a_2 = a_2c_2^3$. If we put $\mathfrak{U}_{21} = \{a_2\}$ and $\mathfrak{U}_{23} = \{c_2\}$, then it is easily seen that both, $\mathfrak{U}_{21}$ and $\mathfrak{U}_{23}$, are of infinite order.[11] If they had an element $\neq 1$ in common, a power a_2^m $(m \neq 0)$ would belong to $\mathfrak{U}_{23}$, and therefore transform c_2 into itself. On the other hand, it transforms c_2 into $c_2^{3^m} \neq c_2$, which is a contradiction. Hence we have again

$$\mathfrak{U}_{21} \cap \mathfrak{U}_{23} = 1.$$

Let $\mathfrak{U}_3$ be the group generated by b_3 and c_3 with the defining relation $(b_3c_3)^2 = 1$. We put $\mathfrak{U}_{31} = \{b_3\}$ and $\mathfrak{U}_{32} = \{c_3\}$.

In order to show that the subgroups $\mathfrak{U}_{ik}$ $(i, k = 1, 2, 3)$ with any of the two possible isomorphisms between the corresponding ones satisfy 3. 0, we need only show, that also $\mathfrak{U}_{31}$ and $\mathfrak{U}_{32}$ are infinite cycles, and $\mathfrak{U}_{31} \cap \mathfrak{U}_{32} = 1$ in $\mathfrak{U}_3$. Hence we have to show that no relation of the form

$$b_3^{\mu}c_3^{\nu} = 1 \text{ with } \mu, \nu \neq 0, 0$$

follows from the defining relation $(b_3c_3)^2 = 1$. To this end, we change the generators of $\mathfrak{U}_3$, *viz.* generate it by c_3 and $b_3c_3 = d$ with the defining relation $d^2 = 1$. Every relation r in $\mathfrak{U}_3$ which follows from this one, is a product of powers of d^2 and its transforms $c_3^{-a}d^2c_3^{a}$. r has, therefore, the property that c_3 occurs in it with a vanishing sum of exponents, and between any two powers of c_3, and possibly at the beginning and the end of r, there stands an even power of d. Now the product $b_3^{\mu}c_3^{\nu}$, expressed in c_3 and d, becomes

$$b_3^{\mu}c_3^{\nu} = (dc_3^{-1})^{\mu}c_3^{\nu}.$$

[11] Using, e. g., the Dehn-Magnus *Freiheitssatz.*

If this equals 1, then it follows from $\mu = 0$ that $\nu = 0$; if $\mu \neq 0$, then we have:

$$b_3{}^{\mu} c_3{}^{\nu} = (dc_3{}^{-1})^{\mu-1} d \cdot c_3{}^{\nu-1} = 1.$$

Here not both, $\mu - 1$ and $\nu - 1$, can be zero. But then the first, i. e., an odd power of d, appears in the beginning of the relation, or between two powers of c. Hence only $\mu = \nu = 0$ is possible.

Now we form the group $\mathfrak{U}$ defined by all the generators and relations of $\mathfrak{U}_1$, $\mathfrak{U}_2$, and $\mathfrak{U}_3$ taken together, and in addition the identifying relations

$$\mathfrak{J}: \quad a_1 = a_2 = a; \qquad b_1 = b_3 = b; \qquad c_2 = c_3 = c.$$

By virtue of the relations $\mathfrak{J}$, $\mathfrak{U}$ is then generated by the elements a, b, c with the defining relations

$$ab = ba, \qquad ca = ac^3, \qquad (bc)^2 = 1.$$

Here the condition (ii) of 1 is not satisfied, i. e., the free product of the groups $\mathfrak{U}_i$ with amalgamated subgroups $\mathfrak{U}_{ik}$ does not exist:

For, the relation $(bc)^2 = 1$ transformed with a becomes:

$$a^{-1}(bc)^2 a = (bc^3)^2 = 1.$$

This, however, is a relation between b and c, which does not follow from $(bc)^2 = 1$. Because, if we express it again by c and $bc = d$, it becomes

$$(bc^3)^2 = (dc^2)^2 = dc^2dc^2 = 1,$$

and here again the first power of d appears between two powers of c.

4. The definitions and theorems of the preceding paragraphs allow the following more general interpretation. This interpretation, though interesting in itself, will not be used in the following paragraphs.

We call the set S of elements $u, v, \cdots$ an incomplete group, if a multiplication is defined for the elements of S, such that

1. for any two elements u, v of S, there exists in S at most one element r such that $uv = r$, at most one element s such that $us = v$, and at most one element t such that $tu = v$;

2. if the products $uv = x$, $vw = y$, $xw = z$, $uy = z'$ exist in S, then $z = z'$; i. e., multiplication, as far as it exists, is associative.

The incomplete group S is called imbeddable, if there exists a group $\mathfrak{H}$ which contains a subsystem $S_{\mathfrak{H}}$ of elements isomorphic with S. We denote the subsystem $S_{\mathfrak{H}}$ of $\mathfrak{H}$ by the same letter S as the original.

The group $\mathfrak{G}$ containing S is said to be freely generated by S, if it is

generated by S, and if every group $\mathfrak{H}$ which contains S and is generated by S, is a homomorphic image of $\mathfrak{G}$, such that the homomorphism from $\mathfrak{G}$ to $\mathfrak{H}$ leaves invariant every element of S.

Now let the system S consist of all the elements of all the groups $\mathfrak{G}_a$, where corresponding elements $u_{\alpha\beta}$ and $u_{\beta\alpha}$ are considered equal as elements of S. Multiplication is defined for any two elements u, v of S which belong to the same group $\mathfrak{G}_a$, i. e. the elements r, s, t of the condition 1 above exist if, and only if, u and v belong both to $\mathfrak{G}_a$, and then r, s, t are the elements uv, $u^{-1}v$, vu^{-1} of $\mathfrak{G}_a$ respectively. If the conditions 3. 0 are fulfilled, S is an incomplete group, and conversely.

3. 2 is an example of an incomplete group consisting of three groups with amalgamated subgroups which is not imbeddable: the element $(bc^3)^2$ of $\mathfrak{U}_3$ which is not the unit element as element of the incomplete group formed by $\mathfrak{U}_1$, $\mathfrak{U}_2$, $\mathfrak{U}_3$, will be reduced to the unit element at any attempt to imbed this incomplete group into a group.

On the other hand, 2. 1 shows, that if this special type of incomplete group is imbeddable, then the generalized free product of the groups $\mathfrak{G}_a$ exists and is freely generated by the incomplete group formed by the $\mathfrak{G}_a$; i. e. if S is imbeddable, then it is imbeddable into a group freely generated by it. This, however, is true for any incomplete group: [12]

4. 0. THEOREM. *If the incomplete group S is imbeddable into a group, then S is imbeddable into a group freely generated by S.*

Proof. Let $\mathfrak{H}$ be any group which contains S. We may assume that $\mathfrak{H}$ is generated by S. If u, v, and w are any three elements of S such that $uv = w$ holds in S, then the relation $uvw^{-1} = 1$ holds in $\mathfrak{H}$.

Now let $\mathfrak{F}$ be the free group generated by all the symbols of S, $\mathfrak{N}$ the smallest self-conjugate subgroup containing all the words of the form uvw^{-1}, where $uv = w$ in S. Then $\mathfrak{H}$ is a homomorphic image of $\mathfrak{F}$,

$$\mathfrak{H} \cong \mathfrak{F}/\mathfrak{M},$$

where $\mathfrak{M}$ is some self-conjugate subgroup of $\mathfrak{F}$. Obviously, $\mathfrak{N}$ belongs to $\mathfrak{M}$. Also, if u and v are any two different symbols of S, they are different also modulo $\mathfrak{M}$. For they are mapped onto different elements of $\mathfrak{H}$. Hence, they are, *a fortiori*, different modulo $\mathfrak{N}$; i. e., $\mathfrak{F}/\mathfrak{N}$ contains S, and the natural homomorphism from $\mathfrak{G} = \mathfrak{F}/\mathfrak{N}$ to $\mathfrak{H} \cong \mathfrak{F}/\mathfrak{M}$ leaves S invariant.

Since $\mathfrak{H}$ is any group which contains, and is generated by, the incomplete group S, $\mathfrak{G}$ is, by definition, freely generated by S.

---

[12] This answers in the affirmative a question raised by R. Baer.

The problem arises to characterize those incomplete groups which can be imbedded into a group; or—in our context—to find sufficient conditions under which an incomplete group consisting of groups with amalgamated subgroups is imbeddable. I have not obtained any general results of this kind. We shall touch upon this problem once more (**9**), after we have gained some more insight into the structure of the generalized free product. But the theorem which follows shows at least that it is sufficient to consider the "reduced" incomplete group formed by the subgroups $\mathfrak{U}_\alpha$ of $\mathfrak{G}_\alpha$: if this incomplete group is imbeddable, then also the incomplete group formed by the groups $\mathfrak{G}_\alpha$ is imbeddable.

5. We are now going to prove the theorem mentioned in the preceding paragraphs:

5.0. THEOREM. *The generalized free product* $\mathfrak{G} = \{\overset{*}{\prod_\alpha} \mathfrak{G}_\alpha; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$ *exists if, and only if, the generalized product* $\mathfrak{U} = \{\overset{*}{\prod_\alpha} \mathfrak{U}_\alpha; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$ *exists.*

Proof. That the condition is necessary, we have seen in **3**. The proof that it is sufficient, consists in the explicit construction of $\mathfrak{G}$. This construction is in principle quite similar to Schreier's construction of the free product with one amalgamated subgroup,[13] but rather more laborious. The procedure is as follows: First we determine in any group $\mathfrak{G}^*$ which is generated by the groups $\mathfrak{G}_\alpha$ such that $\mathfrak{G}_\alpha \cap \mathfrak{G}_\beta = \mathfrak{U}_{\alpha\beta}$ in $\mathfrak{G}^*$, a normal form for the elements of $\mathfrak{G}^*$ in terms of the elements of the groups $\mathfrak{G}_\alpha$. From this we can infer a normal form of the elements of $\mathfrak{G}$, if $\mathfrak{G}$ exists. This knowledge we then use for the construction of $\mathfrak{G}$: We consider all formally different symbols of the same type as this normal form, define a multiplication for them, and show that our symbols form a group with respect to this multiplication. Once this fact is established, it will be easily seen that this group is the generalized free product of the groups $\mathfrak{G}_\alpha$.

Let $\mathfrak{G}^*$ be any group generated by the groups $\mathfrak{G}_\alpha$, such that $\mathfrak{G}_\alpha \cap \mathfrak{G}_\beta = \mathfrak{U}_{\alpha\beta}$ in $\mathfrak{G}^*$. Then every element G^* of $\mathfrak{G}^*$ is a product of elements G_α of the groups $\mathfrak{G}_\alpha$:

5.01. $$G^* = \prod_{\nu=1}^{n} G_{\alpha_\nu}, \text{ where } G_{\alpha_\nu} \neq 1 \text{ for } \nu = 1, \cdots, n,$$
$$\text{and } \alpha_\nu \neq \alpha_{\nu+1} \text{ for } \nu = 1, \cdots, n-1.$$

We denote by $\mathfrak{U}^*$ the subgroup of $\mathfrak{G}^*$ which is generated by the subgroups $\mathfrak{U}_\alpha$ of $\mathfrak{G}_\alpha$. In $\mathfrak{U}^*$ we choose a system of right-hand representatives with respect to every one of its subgroups $\mathfrak{U}_\alpha$; let it consist of the elements $\bar{U}_0{}^\alpha = 1$, $\bar{U}_\kappa{}^\alpha$,

[13] Cf. [6].

where κ varies in a certain set of suffixes depending on α. We also choose in every group $\mathfrak{G}_\alpha$ a system of right-hand representatives with respect to its subgroup $\mathfrak{U}_\alpha$, which we denote by $F_0{}^\alpha = 1$, $F_\lambda{}^\alpha$. Finally, for later use, we choose also in $\mathfrak{U} = \{\overset{*}{\prod_\alpha} \mathfrak{U}_\alpha; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$ a system of right-hand representatives with respect to every one of its subgroups $\mathfrak{U}_\alpha$, *viz.* $U_0{}^\alpha = 1$, $U_\mu{}^\alpha$.

Now we replace in 5.01 every factor G_{α_ν} by its representative $F_\lambda{}^{\alpha_\nu}$ by means of:

$$G_{\alpha_\nu} = U_{\alpha_\nu} F_\lambda{}^{\alpha_\nu} \qquad (U_{\alpha_\nu} \text{ in } \mathfrak{U}_{\alpha_\nu}).$$

If $F_\lambda{}^{\alpha_\nu} = 1$, we combine U_{α_ν} and $U_{\alpha_{\nu+1}}$ into one element $U^*{}_\nu$ of $\mathfrak{U}^*$. After this, 5.01 will assume the form

5.02. $$G^* = U^*{}_0 \prod_{\mu=1}^{m} (F^{\alpha_\mu} U^*{}_\mu),$$

where $U^*{}_\mu$ in $\mathfrak{U}^*$ and $F^{\alpha_\mu} \neq 1$ for all μ.

Now, starting from the right, we replace step by step every $U^*{}_\mu$ by its representative with respect to $\mathfrak{U}_{\alpha_\mu}$:

$$U^*{}_\mu = U_{\alpha_\mu} \bar{U}_\kappa{}^{\alpha_\mu} \qquad (U_{\alpha_\mu} \text{ in } \mathfrak{U}_{\alpha_\mu}),$$

then replace $F_\lambda{}^{\alpha_\mu} U_{\alpha_\mu}$ by

$$F_\lambda{}^{\alpha_\mu} U_{\alpha_\mu} = U'{}_{\alpha_\mu} F_{\lambda'}{}^{\alpha_\mu},$$

and combine $U'{}_{\alpha_\mu}$ with $U^*{}_{\mu-1}$. After this process has been applied m times, 5.02 will be of the form

5.03. $$G^* = U^* \prod_{\mu=1}^{m} (F^{\alpha_\mu} \bar{U}_\kappa{}^{\alpha_\mu}) \qquad (U^* \text{ in } \mathfrak{U}^*),$$

where we may assume that $\alpha_\mu \neq \alpha_{\mu+1}$ if $\bar{U}_\kappa{}^{\alpha_\mu} = 1$; for otherwise F^{α_μ} and $F^{\alpha_{\mu+1}}$ can be combined into one factor.

5.03 represents the desired normal form for the elements of $\mathfrak{G}^*$.

Before we go on, we fix the notation used in this paragraph: $\mathfrak{U}$ always denotes the free product of the groups $\mathfrak{U}_\alpha$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$; its elements will be $U, V, W, U', \cdots$; the elements of $\mathfrak{U}_\alpha$ will be $U_\alpha, U'_\alpha, \cdots$, and the right-hand representatives of $\mathfrak{U}$ with respect to $\mathfrak{U}_\alpha$ are $U_0{}^\alpha = 1$, $U_\mu{}^\alpha$. Right-hand representatives of $\mathfrak{G}_\alpha$ with respect to $\mathfrak{U}_\alpha$ are, as before, $F_0{}^\alpha = 1$, $F_\lambda{}^\alpha$; and arbitrary elements of $\mathfrak{G}_\alpha$ are denoted by $G_\alpha, G'_\alpha, \cdots$.

Now we consider all the different symbols of the form 5.03, but with the generalized free product $\mathfrak{U}$ in place of $\mathfrak{U}^*$; i.e., all symbols of the form

5.1. $$R = U_0 L_1 U_1 \cdots L_r U_r, \qquad r \geq 0,$$

where 1. $L_\rho = F_\lambda^{\alpha_\rho} \neq 1$ for $\rho = 1, \cdots, r$,

2. U_0 any element of $\mathfrak{U}$,

3. $U_\rho = U_\mu^{\alpha_\rho}$ for $\rho = 1, \cdots, r$,

4. If $U_\rho = 1$, then $\alpha_\rho \neq \alpha_{\rho+1}$ $(1 \leq \rho \leq r-1)$.

$r = l(R)$ is called the length of R. Two symbols 5.1 are considered equal, if, and only if, they are identical.

Now we define the product of two symbols

$$R = U_0 L_1 U_1 \cdots L_r U_r$$

and

$$S = V_0 M_1 V_1 \cdots M_s V_s$$

in the following way:

D0. If $r = 0$: $RS = (U_0 V_0) M_1 V_1 \cdots M_s V_s$, (in particular, the multiplication of two symbols of length zero is defined as that in $\mathfrak{U}$).

D1. 1. If $r = 1$, and if L_1, $(U_1 V_0)$, and M_1 all belong to the same group $\mathfrak{G}_\alpha$ (i. e. $U_1 V_0$ belongs to the subgroup $\mathfrak{U}_\alpha$ of $\mathfrak{U}$), and if

$$L_1 (U_0 V_0) M_1 = U_\alpha F_\lambda^\alpha \text{ in } \mathfrak{G}_\alpha,$$

then we define

$$RS = (U_0 U_\alpha) F_\lambda^\alpha V_1 M_2 V_2 \cdots M_s V_s, \text{ if } F_\lambda^\alpha \neq 1,$$

and

$$RS = (U_0 U_\alpha V_1) M_2 V_2 \cdots M_s V_s, \text{ if } F_\lambda^\alpha = 1.$$

D1. 2. If $r = 1$, and if L_1, $(U_1 V_0)$, and M_1 do not all belong to the same group $\mathfrak{G}_\alpha$ as L_1; and if

$$\text{in } \mathfrak{U}\colon \quad U_1 V_0 = U_\alpha U_\mu^\alpha,$$

$$\text{and in } \mathfrak{G}_\alpha\colon \quad L_1 U_\alpha = U'_\alpha F_\lambda^\alpha,$$

then we define

$$RS = (U_0 U'_\alpha) F_\lambda^\alpha U_\mu^\alpha M_1 V_1 \cdots M_s V_s.$$

D2. Let us assume the product RS to be defined for all symbols R of length $\leq r-1$. If R is a symbol of length r, $R = U_0 L_1 \cdots L_{r-1} U_{r-1} L_r U_r$, we denote by R^* the symbol

$$R^* = U_0 L_1 \cdots L_{r-1} U_{r-1}.$$

Then we define

$$RS = R^* \cdot [(L_r U_r) \cdot S].$$

It is immediately clear, that with this definition of multiplication, the product of any two symbols is again a symbol 5.1 with the properties 1.-4. In order to prove that this multiplication defines a group, we have to show that it is associative, that there exists a unit element, and that every symbol possesses an inverse. The proof of the associative law presents the main difficulty. We shall need for it the following three lemmas:

5.2. *If R and S are any two symbols with the property that written in conjunction, $\{R,S\} = U_0L_1 \cdots L_rU_rV_0M_1 \cdots M_sV_s$, they represent again a symbol* 5.1, *then this symbol $\{R,S\}$ is the product of R and S.*

$$\{R,S\} = R \cdot S.$$

In particular, every symbol R of length $r \geq 1$, is the product of $R^* = U_0L_1 \cdots L_{r-1}U_{r-1}$ and L_rU_r: $R = R^* \cdot (L_rU_r)$.

Proof. For symbols R of length $0, 1$ and arbitrary S, this follows immediately from D0 and D1.2 respectively. Let it be true for all symbols R of length $l(R) \leq r - 1$ and arbitrary S. We prove it for symbols R of length r and arbitrary S:

Since R and S fit together to form one symbol $\{R,S\}$, the same is true of the last term L_rU_r of R, and S, and as L_rU_r is of length one, we know that $\{L_rU_r, S\} = (L_rU_r) \cdot S$. But the symbol $\{L_rU_r, S\}$ begins with L_rU_r, so that also $R^* = U_0L_1 \cdots L_{r-1}U_{r-1}$ and $(L_rU_r) \cdot S$ satisfy the assumption of the lemma. But $l(R^*) = r - 1$, so that by induction, the symbol $\{R^*, (L_rU_r) \cdot S\}$ is the product of R^* and $(L_rU_r) \cdot S$:

$$\{R^*, (L_rU_r) \cdot S\} = R^* \cdot [(L_rU_r) \cdot S].$$

By D2, the right-hand side is the product $R \cdot S$, while the symbol on the left-hand side is the same as the symbol $\{R,S\}$. Hence $\{R,S\} = R \cdot S$.

5.3. *If R and S are any two symbols, then*

$$l(RS) \leq l(R) + l(S).$$

Proof. For $l(R) = 0, 1$, 5.3 is obvious from D0 and D1 respectively. Let it be true for $l(R) \leq r - 1$, and let R be a symbol of length r. Then

$$R = R^* \cdot L_rU_r \text{ with } l(R^*) = r - 1,$$

and by D2:

$$RS = R^*[(L_rU_r) \cdot S].$$

Hence, by induction:

$$\begin{aligned} l(RS) &\leq l(R^*) + l[(L_rU_r) \cdot S] \\ &\leq l(R^*) + l(L_rU_r) + l(S) \\ &= l(R) + l(S). \end{aligned}$$

5.4. *If R and S are any two symbols, and $S = V_0M_1V_1 \cdots M_sV_s$ where $s \geq 1$, then either $l(RS) < l(R)$, or the product RS ends on the terms M'_sV_s, where M'_s belongs to the same group $\mathfrak{G}_a$ as M_s.*

Proof. The lemma is obvious for $l(R) = 0$, by D0. If $l(R) = 1$, and if D1.2 applies, then of the whole symbol S only the term V_0 is affected in the product RS, so that 5.4 is true in this case. If D1.1 applies, certainly the part $M_rV_r \cdots M_sV_s$ of S is unaffected in the product RS; which proves 5.4 in case $s \geq 2$. But if $s = 1$, we have

$$RS = (U_0L_1U_1) \cdot (V_0M_1V_1) = U'_0M'_1V_1,$$

where M'_1 is either a representative $\neq 1$ out of the same group as M_1, or $G'_1 = 1$, and then $l(RS) = 0 < l(R)$.

Now let 5.4 be true for $l(R) \leq r - 1$, and let R be a symbol of length r. Then

$$R = R^* \cdot L_rU_r \text{ with } l(R^*) = r - 1,$$

and

$$RS = R^* \cdot [(L_rU_r) \cdot S] = R^* \cdot S',$$

with

$$S' = (L_rU_r) \cdot S.$$

Hence, by induction, either $l(S') < l(L_rU_r)$, i. e. $l(S') = 0$, and then, by 5.3:

$$l(RS) \leq l(R^*) + l(S') = r - 1 < r;$$

or S' ends on a term M''_sV_s, where M''_s belongs to the same group as M_s. Besides, as $l(R^*) = r - 1$, our lemma holds for the product R^*S'. Hence either

$$l(RS) = l(R^*S') < l(R^*) = r - 1 < r,$$

or $RS = R^*S'$ ends on a term M'_sV_s, where M'_s belongs to the same group as M''_s, i. e. to the same group as M_s. Which proves the lemma.

We come now to the proof of the associative law. Let

$$\begin{aligned} R &= U_0L_0U_1 \cdots L_\rho U_\rho, \\ S &= V_0M_0V_1 \cdots M_\sigma V_\sigma, \\ T &= W_0N_0W_1 \cdots N_\tau W_\tau. \end{aligned}$$

be three symbols of length ρ, σ, τ respectively. We have to show that the multiplications defined by D0-D2 satisfies

5.5. $$(RS)T = R(ST).$$

Proof. We prove 5.5 by an induction with respect to ρ. To this end we

note first that the defining equation D2 holds also for $r = 1$, as is obvious from the definition D1. Hence any induction with respect to the length which is based on the inductive definition D2, may be started with length zero if otherwise convenient.

1. $\rho = 0$.

We prove this case by an induction with respect to σ for arbitrary τ. If $\sigma = 0$, we have by D0:

$$(RS)T = (U_0V_0)(W_0N_1W_1 \cdots N_\tau W_\tau) = (U_0V_0W_0)N_1W_1 \cdots N_\tau W_\tau,$$

and

$$R(ST) = U_0[(V_0W_0)N_1W_1 \cdots N_\tau W_\tau) = (U_0V_0W_0)N_1W_1 \cdots N_\tau W_\tau;$$

hence $(RS)T = R(ST)$.

Let 5. 5 be true for $\rho = 0$, $\sigma \leqq s - 1$, τ arbitrary; then it is also true for $\rho = 0$, $\sigma = s$, τ arbitrary. For, let S be a symbol of length $s > 0$; then

$$S = S^*M_sV_s \text{ with } l(S^*) = s - 1 \geqq 0.$$

By D0.: $\quad (RS)T = [(U_0V_0)M_1V_1 \cdots M_sV_s] \cdot T$

by D2.: $\quad = [(U_0V_0)M_1V_1 \cdots M_{s-1}V_{s-1}] \cdot [(M_sV_s) \cdot T]$

by D0.: $\quad = (U_0 \cdot S^*) \cdot [(M_sV_s) \cdot T]$.

And by D2.: $\quad R(ST) = U_0 \cdot [S^* \cdot (M_sV_s \cdot T)]$

by induction: $\quad = (U_0S^*)[(M_sV_s) \cdot T]$.

Hence $R(ST) = (RS)T$, i. e. 5. 5 holds for $\rho = 0$ and arbitrary σ and τ.

2. We assume 5. 5 to be true for all symbols R of length $\rho \leqq r - 1$ and symbols S and T of arbitrary length σ and τ respectively. Then we have to prove it for symbols R of length r and symbols S and T of arbitrary length σ and τ respectively. This we do again by an induction with respect to σ.

a. $\sigma = 0$.

We have: $R = R^*L_rU_r$, $l(R^*) = r - 1$; and $S = V_0$.

Hence by D2.: $\quad (RS)T = [R^* \cdot (L_rU_r \cdot S)] \cdot T$

by induction: $\quad = R^* \cdot [(L_rU_r \cdot S) \cdot T] = R^* \cdot [(L_rU_r \cdot V_0) \cdot T]$.

And by D0.: $\quad R(ST) = R \cdot [(V_0W_0) \cdot N_1W_1 \cdots N_\tau W_\tau]$

by D2.: $\quad = R^* \cdot \{(L_rU_r) \cdot [(V_0W_0)N_1W_1 \cdots N_\tau W_\tau]\}$.

It remains to be shown that

$$(L_rU_r \cdot V_0)(W_0N_1W_1 \cdots N_\tau W_\tau) \\ = (L_rU_r)[(V_0W_0)N_1W_1 \cdots N_\tau W_\tau].$$

Here each side has to be calculated according to D1. If L_r belongs to $\mathfrak{G}_a$, then on either side the first step of this calculation amounts to representing the element $(U_rV_0)W_0 = U_r(V_0W_0)$ of $\mathfrak{U}$ in the form $U_aU_\mu{}^a$; and the remaining steps are then identical on both sides. Hence both sides are equal.

b. We have to deal with the case $\sigma = 1$ separately, i. e. we have to prove 5. 5 for $\rho = r$, $\sigma = 1$ and arbitrary τ under the assumption that it is true for $\rho \leq r - 1$ and all σ and τ, and for $\rho = r$, $\sigma = 0$, and arbitrary τ.

We have

$$R = R^*L_rU_r, \qquad l(R^*) = r - 1,$$

and

$$S = V_0M_1V_1.$$

By D2.: $\quad (RS)T = [R^* \cdot (L_rU_r \cdot S)]T$

by induction: $\quad = R^*[(L_rU_r \cdot S)T]$.

And by D2.: $\quad R(ST) = R^*[L_rU_r \cdot (ST)]$
$\quad = R^*[L_rU_r(V_0M_1V_1 \cdot T)]$

by D2.: $\quad = R^*[L_rU_r(V_0 \cdot (M_1V_1 \cdot T))]$

by a.: $\quad = R^*[(L_rU_r \cdot V_0)(M_1V_1 \cdot T)]$.

Hence it remains to be shown that

$$(L_rU_r \cdot V_0M_1V_1)(W_0N_1W_1 \cdots N_\tau W_\tau) \\ = (L_rU_r \cdot V_0)(M_1V_1 \cdot W_0N_1W_1 \cdots N_\tau W_\tau).$$

On either side, we have to form first (by D1):

$$L_rU_r \cdot V_0 = \bar{U}_{r-1}\bar{L}_r\bar{U}_r,$$

where $\bar{L}_r = \bar{F}_\lambda{}^a$ is a representative out of the same group $\mathfrak{G}_a$ as L_r, and $\bar{U}_r = \bar{U}_\mu{}^a$ a representative of $\mathfrak{U}$ with respect to $\mathfrak{U}_a$.

If $\bar{L}_r$, $\bar{U}_r$, and M_1 do not all belong to $\mathfrak{G}_a$, then

$$L_rU_r \cdot V_0M_1V_1 = \bar{U}_{r-1}\bar{L}_r\bar{U}_rM_1V_1$$

is itself one of our symbols 5. 1, and both sides are equal by 5. 2 and D2.

If $\bar{L}_r$, $\bar{U}_r$, and M_1 all belong to $\mathfrak{G}_a$ (i. e., $\bar{U}_r = 1$), and if (V_1W_0) and N_1 also belong to $\mathfrak{G}_a$, then again both sides are equal because of the associative law in $\mathfrak{G}_a$.

It remains the case, that $\bar{L}_r$, $\bar{U}_r = 1$, and M_1 all belong to $\mathfrak{G}_\alpha$, but $(V_1 W_0)$ and N_1 do not both belong to $\mathfrak{G}_\alpha$. In that case we have to prove:

$$[\bar{U}_{r-1}(\bar{L}_r M_1) \cdot V_1](W_0 N_1 W_1 \cdots N_\tau W_\tau)$$
$$= (\bar{U}_{r-1}\bar{L}_r)(M_1 V_1 \cdot W_0 N_1 W_1 \cdots N_\tau W_\tau).$$

By D1., we have to form on the left-hand side

$$V_1 W_0 = U_\alpha U_\mu{}^\alpha,$$

and then

$$(\bar{L}_r M_1) U_\alpha = \bar{U}_\alpha \bar{M}_1 \qquad (\bar{M}_1 \text{ in } \mathfrak{G}_\alpha).$$

Then the left-hand side becomes:

$$(\bar{U}_{r-1}\bar{U}_\alpha)\bar{M}_1 U_\mu{}^\alpha N_1 W_1 \cdots N_\tau W_\tau.$$

On the right-hand side, we also have to form $V_1 W_0 = U_\alpha U_\mu{}^\alpha$, but then we have to normalize $M_1 U_\alpha$ first, and only then to multiply $\bar{L}_r$ by it, i. e., we have to form $\bar{L}_r(M_1 U_\alpha)$. But since in $\mathfrak{G}_\alpha$ the associative law holds, we have

$$\bar{L}_r(M_1 U_\alpha) = (\bar{L}_r M_1) U_\alpha = \bar{U}_\alpha \bar{M}_1,$$

so that the right-hand side leads to the same result

$$(\bar{U}_{r-1}\bar{U}_\alpha)\bar{M}_1 U_\mu{}^\alpha N_1 W_1 \cdots N_\tau W_\tau.$$

c. Now let us assume 5. 5 to be true for $\rho = r$, $\sigma \leq s - 1$, and arbitrary τ; then we prove it for $\rho = r$, $\sigma = s$, and arbitrary τ. As we may now assume that $s > 1$, we have

$$R = R^* L_r U_r, \qquad l(R^*) = r - 1,$$

and

$$S = S^* M_s V_s, \qquad l(S^*) = s - 1 \geq 1.$$

Now $\qquad (RS)T = [R(S^* M_s V_s)] \cdot T$

by 5. 2 and induction: $\qquad = [(RS^*) M_s V_s] \cdot T.$

And by D2.: $\qquad R(ST) = R[S^*(M_s V_s \cdot T)]$

by induction: $\qquad = (RS^*)(M_s V_s \cdot T).$

Now we apply 5. 4, using $l(S^*) \geq 1$. Thereby we have either

$$l(RS^*) < l(R) = r,$$

in which case by induction:

$$[(RS^*) M_s V_s] T = (RS^*)(M_s V_s \cdot T).$$

Or else, RS^* ends on a term $M'_{s-1} V_{s-1}$, where M'_{s-1} belongs to the same group

$\mathfrak{G}_\alpha$ as M_{s-1}. But then $(RS^*)M_sV_s$ is, as it stands, a symbol 5.1, and therefore by D2.:

$$[(RS^*)M_sV_s]T = (RS^*)(M_sV_s \cdot T).$$

This completes the proof of the associative law.

From D0. it follows: the symbol of length zero for which U_0 is the unit element of $\mathfrak{U}$, is a unit element for the multiplication defined by D0.-D2.

The symbols of length one which are of the form $U_\alpha F_\mu{}^\alpha$, correspond exactly to all the elements of the group $\mathfrak{G}_\alpha$ for every α. By D0. and D1., their multiplication is the same as in $\mathfrak{G}_\alpha$. Every group $\mathfrak{G}_\alpha$ is, therefore, isomorphically represented in the system $\mathfrak{G}$ of all symbols 5.1.

Every one of these symbols which correspond to elements of the groups $\mathfrak{G}_\alpha$, possesses an inverse in the system $\mathfrak{G}$, *viz.* the symbol corresponding to its inverse in $\mathfrak{G}_\alpha$. On the other hand, all the symbols of $\mathfrak{G}$ are products of symbols corresponding to elements of the groups $\mathfrak{G}_\alpha$. But then it follows from the associative law that every symbol 5.1 possesses an inverse. Hence: the system $\mathfrak{G}$ of all formally different symbols 5.1 forms a group with respect to the multiplication D0.-D2.

If we denote the subsystems consisting of all the symbols $U_\alpha F_\lambda{}^\alpha$ (α fixed) again by $\mathfrak{G}_\alpha$, then $\mathfrak{G}$ is generated by its subgroups $\mathfrak{G}_\alpha$. Moreover, $\mathfrak{G}_\alpha \cap \mathfrak{G}_\beta = \mathfrak{U}_{\alpha\beta}$ in $\mathfrak{G}$. For two symbols 5.1 are equal only if they are identical. Hence

$$U_\alpha F_\lambda{}^\alpha = U_\beta F_{\lambda'}{}^\beta, \qquad \alpha \neq \beta$$

implies $F_\lambda{}^\alpha = F_{\lambda'}{}^\beta = 1$, and then $U_\alpha = U_\beta = U_{\alpha\beta}$, since the meet in $\mathfrak{U}$ of $\mathfrak{U}_\alpha$ and $\mathfrak{U}_\beta$ is $\mathfrak{U}_{\alpha\beta}$.

It follows that—with the notations of 1—the relations $\mathfrak{R}_\alpha$ for all α, and the relations $\mathfrak{J}$, hold in $\mathfrak{G}$. But every other relation $R = 1$ follows from these. For, if we represent R in the form 5.1, R is of length zero, i. e. an element of $\mathfrak{U}$. And as $\mathfrak{U} = \{\overset{*}{\prod_\alpha} \mathfrak{U}_\alpha ; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$, every relation in $\mathfrak{U}$ follows from those in $\mathfrak{U}_\alpha$ and from the relations of $\mathfrak{J}$.

This completes the proof of Theorem 5.0.

Incidentally, our construction has shown:

5.6. THEOREM. *The subgroups $\mathfrak{U}_\alpha$ of $\mathfrak{G}_\alpha$ generate in $\mathfrak{G}$ the generalized free product $\mathfrak{U}$.*

For we constructed the group $\mathfrak{G}$ so that it contains $\mathfrak{U}$.

Also, we saw that in any group $\mathfrak{G}^*$ which is generated by the groups $\mathfrak{G}_\alpha$

such that $\mathfrak{G}_\alpha \cap \mathfrak{G}_\beta = \mathfrak{U}_{\alpha\beta}$ in $\mathfrak{G}^*$, the elements can be represented in the normal form 5.03 corresponding to the normal form 5.1 in $\mathfrak{G}$. We know now:

5.7. THEOREM. *$\mathfrak{G}^*$ is the generalized free product $\mathfrak{G}$ if, and only if, the subgroups $\mathfrak{U}_\alpha$ of $\mathfrak{G}_\alpha$ generate the generalized free product $\mathfrak{U}$ in $\mathfrak{G}^*$, and different normal forms* 5.03 (*i. e.* 5.1, *as* $\mathfrak{U}^* = \mathfrak{U}$) *represent different elements of* $\mathfrak{G}^*$.

6. In this paragraph we describe the structure of the generalized free product $\mathfrak{G}$ in a different way, which will prove useful later on.

For every suffix α, we denote by $\mathfrak{G}^\alpha$ the subgroup of $\mathfrak{G}$ which is generated by all the groups $\mathfrak{G}_\beta$ $(\beta \neq \alpha)$. Then $\mathfrak{G}$ is generated by $\mathfrak{G}_\alpha$ and $\mathfrak{G}^\alpha$.

The elements of $\mathfrak{G}_\alpha$ in $\mathfrak{G}$ are exactly those whose normal form 5.1 is of the form U_0L_1 (U_0 in $\mathfrak{U}_\alpha$, $L_1 = F_\lambda{}^\alpha$ in $\mathfrak{G}_\alpha$). Since $\mathfrak{G}^\alpha$ is generated by all the groups $\mathfrak{G}_\beta$ with $\beta \neq \alpha$, the normal form of an element of $\mathfrak{G}^\alpha$ does not contain any representatives $F_\lambda{}^\alpha$ of $\mathfrak{G}_\alpha$. Hence the meet of $\mathfrak{G}^\alpha$ and $\mathfrak{G}_\alpha$ belongs to $\mathfrak{U}$. But the whole group $\mathfrak{U}$ belongs to $\mathfrak{G}^\alpha$; for all groups $\mathfrak{U}_\beta$ $(\beta \neq \alpha)$ belong to $\mathfrak{G}^\alpha$, and the group $\mathfrak{U}_\alpha$ also does, as it is generated by all the groups $\mathfrak{U}_{\alpha\beta}$ (α fixed) which, because of $\mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}$ in $\mathfrak{G}$, are also subgroups of the groups $\mathfrak{G}_\beta$ $(\beta \neq \alpha)$. Therefore, we have

6.01 $$\mathfrak{G}_\alpha \cap \mathfrak{G}^\alpha = \mathfrak{G}_\alpha \cap \mathfrak{U} = \mathfrak{U}_\alpha \text{ for every } \alpha.$$

However, we can show more, *viz.*

6.02. THEOREM. *$\mathfrak{G}$ is the free product of the groups $\mathfrak{G}_\alpha$ and $\mathfrak{G}^\alpha$ with the amalgamated subgroup $\mathfrak{U}_\alpha$,*

$$\mathfrak{G} = \mathfrak{G}_\alpha \underset{\mathfrak{U}_\alpha}{*} \mathfrak{G}^\alpha \textit{ for every } \alpha.$$

Proof. Let us denote the free product of $\mathfrak{G}_\alpha$ and $\mathfrak{G}^\alpha$ with the amalgamated subgroup $\mathfrak{U}_\alpha$ by $\mathfrak{F}$:

$$\mathfrak{F} = \mathfrak{G}_\alpha \underset{\mathfrak{U}_\alpha}{*} \mathfrak{G}^\alpha.$$

As $\mathfrak{G}$ is generated by $\mathfrak{G}_\alpha$ and $\mathfrak{G}^\alpha$, and the meet of these two groups is $\mathfrak{U}_\alpha$ in $\mathfrak{G}$, we have by 2.1:

$$\mathfrak{G} \cong \mathfrak{F}/\mathfrak{N},$$

where $\mathfrak{N}$ is a self-conjugate subgroup of $\mathfrak{F}$ which is tidy with respect to $\mathfrak{G}_\alpha$ and $\mathfrak{G}^\alpha$. On the other hand, $\mathfrak{F}$ also is generated by *all* the groups $\mathfrak{G}_\alpha$, and any two of them have meet $\mathfrak{U}_{\alpha\beta}$ in $\mathfrak{F}$. Hence, again by 2.1, $\mathfrak{F}$ itself is a homomorphic image of $\mathfrak{G}$, such that the subgroups $\mathfrak{G}_\alpha$ of $\mathfrak{F}$ are the maps of the subgroups $\mathfrak{G}_\alpha$ of $\mathfrak{G}$. It follows:

$$\mathfrak{G} \cong \mathfrak{F}.$$

In order to determine the structure of $\mathfrak{G}^{\alpha}$, we write the whole system of generators and relations of $\mathfrak{G}$ (*cf.* 1) arranged so as to express 6.02 in terms of generators and relations:

a. We begin with the generators $u_{\alpha\beta}$ (α fixed) and g_{α} of $\mathfrak{G}_{\alpha}$. The system $\mathfrak{R}_{\alpha}$ of defining relations for these generators we choose so that it contains a subsystem $\mathfrak{r}_{\alpha}$ of defining relations for the generators $u_{\alpha\beta}$ of $\mathfrak{U}_{\alpha}$—i.e., the generators $u_{\alpha\beta}$ with the relations of $\mathfrak{r}_{\alpha}$ define $\mathfrak{U}_{\alpha}$.

b. Then we take all generators $u_{\beta\gamma}$ and g_{β} of all the groups $\mathfrak{G}_{\beta}$ $(\beta \neq \alpha)$ with the system of defining relations consisting of

1. all the systems $\mathfrak{R}_{\beta}$ $(\beta \neq \alpha)$,

2. all those identifying relations $u_{\beta\gamma} = \mathfrak{J}_{\beta\gamma}(u_{\beta\gamma})$, where $\beta \neq \alpha \neq \gamma$,

3. the system $\mathfrak{r}^{\alpha}$ of relations which is obtained from $\mathfrak{r}_{\alpha}$ by replacing every generator $u_{\alpha\beta}$ by $u_{\beta\alpha} = \mathfrak{J}_{\alpha\beta}(u_{\alpha\beta})$. (These relations certainly hold in $\mathfrak{G}$—they follow from the systems $\mathfrak{r}_{\alpha}$ and $\mathfrak{J}$—, so that we are allowed to add them to the relations of $\mathfrak{G}$.)

The system of generators $u_{\beta\alpha}$ (α fixed) with the defining relations $\mathfrak{r}^{\alpha}$ defines a group $\mathfrak{U}^{\alpha}$ which is isomorphic with $\mathfrak{U}_{\alpha}$.

c. Finally, we add the remaining identifying relations $u_{\alpha\beta} = \mathfrak{J}_{\alpha\beta}(u_{\alpha\beta})$ for all elements $u_{\alpha\beta}$ of the subgroups $\mathfrak{U}_{\alpha\beta}$ (α fixed) of $\mathfrak{G}_{\alpha}$.

These last relations identify the subgroup $\mathfrak{U}_{\alpha}$ of $\mathfrak{G}_{\alpha}$ with the subgroup $\mathfrak{U}^{\alpha}$ of the group defined by *b*. Since a., b., and c. together define exactly the group $\mathfrak{G}$, the group defined by b. must be $\mathfrak{G}^{\alpha}$: for, $\mathfrak{G}^{\alpha}$ is certainly generated by the generators of all the groups $\mathfrak{G}_{\beta}$ $(\beta \neq \alpha)$, and the relations b. 1.-3. hold in $\mathfrak{G}^{\alpha}$; if there were relations in $\mathfrak{G}^{\alpha}$ which do not follow from these, then they certainly follow after the remaining defining relations of $\mathfrak{G}$ have been added, i.e. the relations of a. defining $\mathfrak{G}_{\alpha}$, and the relations of c. But this would be a contradiction to 6.02.

Since $\mathfrak{G}^{\alpha}$ can be defined by the generators and relations given under b., it has the following structure:
Let

$$\mathfrak{F}^{\alpha} = \{\prod_{\beta \neq \alpha}^{*} \mathfrak{G}_{\beta}; \mathfrak{U}_{\beta\gamma} = \mathfrak{U}_{\gamma\beta}\}$$

—$\mathfrak{F}^{\alpha}$ exists by 2.1—, let $\bar{\mathfrak{U}}^{\alpha}$ be the subgroup of $\mathfrak{F}^{\alpha}$ generated by all the subgroups $\mathfrak{U}_{\beta\alpha}$ (α fixed) of $\mathfrak{F}^{\alpha}$. Let $\mathfrak{N}^{\alpha}$ be the smallest self-conjugate subgroup of $\mathfrak{F}^{\alpha}$ which contains the left-hand sides of the relations of the system $\mathfrak{r}^{\alpha}$. Then

6.1. $\mathfrak{G}^a \cong \mathfrak{F}^a/\mathfrak{N}^a$ with $\bar{\mathfrak{U}}^a/\bar{\mathfrak{U}}^a \cap \mathfrak{N}^a \cong \mathfrak{U}^a \cong \mathfrak{U}_a$.

6.02 and 6.1 together obviously characterize $\mathfrak{G}$ completely:

6.20. Criterion. *$\mathfrak{G}$ is the free product of the groups $\mathfrak{G}_a$ with amalgamated subgroups $\mathfrak{U}_{a\beta}$ if, and only if, for some suffix α,*

$$\mathfrak{G} = \mathfrak{G}_a \underset{\mathfrak{U}_a}{*} \mathfrak{G}^a,$$

where $\mathfrak{G}^a$ is isomorphic to the factor group of the generalized free product $\mathfrak{F}^a$ of the groups $\mathfrak{G}_\beta$ $(\beta \neq \alpha)$ with respect to a tidy self-conjugate subgroup $\mathfrak{N}^a$ of $\mathfrak{F}^a$ which can be generated by elements of the subgroup $\bar{\mathfrak{U}}^a$ and their transforms in $\mathfrak{F}^a$, and which has the property that

$$\bar{\mathfrak{U}}^a/\bar{\mathfrak{U}}^a \cap \mathfrak{N}^a \cong \mathfrak{U}_a.$$

In the special case of the free product with one amalgamated subgroup this amounts to:

6.21. Criterion. *$\mathfrak{G}$ is the free product of the groups $\mathfrak{G}_a$ with one amalgamated subgroup $\mathfrak{U}$ if, and only if, for some suffix α,*

$$\mathfrak{G} = \mathfrak{G}_a \underset{\mathfrak{U}}{*} \mathfrak{G}^a, \text{ where } \mathfrak{G}^a = \{\overset{*}{\prod_{\beta\neq a}} \mathfrak{G}_\beta ; \mathfrak{U} = \mathfrak{U}\}.$$

For a finite number of factors $\mathfrak{G}_a$, this leads to a different and rather more useful criterion, as we shall see in the next paragraph.

7. We have seen in **5** that in the generalized free product $\mathfrak{G}$ of the groups $\mathfrak{G}_a$, the subgroups $\mathfrak{U}_a$ of $\mathfrak{G}_a$ generate their generalized free product $\mathfrak{U}$. This suggests that the same might be true for every system of groups $\mathfrak{H}_a$ "between" $\mathfrak{U}_a$ and $\mathfrak{G}_a$; i. e., subgroups of $\mathfrak{G}_a$ containing $\mathfrak{U}_a$. This is, in fact, true, and is the generalization of the fact that in the ordinary free product $\overset{*}{\prod_a} \mathfrak{G}_a$ subgroups of the factors $\mathfrak{G}_a$ generate a group which is the free product of these subgroups.

The fact that the groups $\mathfrak{U}_a$ generate the generalized free product $\mathfrak{U}$ in $\mathfrak{G}$ we proved by an indirect method: we constructed a group containing $\mathfrak{U}$, and proved afterwards that this group was the generalized free product $\mathfrak{G}$.

For the proof of the corresponding fact for the subgroups $\mathfrak{H}_a$ of $\mathfrak{G}_a$ we use the Theorem 5.7, which is easily applicable in this case, as the normal form for the elements in $\mathfrak{H}$ can be chosen so that it is at the same time a normal form for these elements considered as elements of $\mathfrak{G}$. But

5.7 will also help us in other cases, in particular where we are dealing with free products with one amalgamated subgroup where the first part of 5.7, that the subgroups $\mathfrak{U}_a$ generate the free product $\mathfrak{U}$, is trivially satisfied: all $\mathfrak{U}_a$ are identical with $\mathfrak{U}$.

We are now going to prove:

7.0. THEOREM. *If* $\mathfrak{G} = \{\overset{*}{\prod_a} \mathfrak{G}_a ; \mathfrak{U}_{a\beta} = \mathfrak{U}_{\beta a}\}$ *and if, for every* α, $\mathfrak{H}_a$ *is a subgroup of* $\mathfrak{G}_a$ *containing* $\mathfrak{U}_a$, $\mathfrak{U}_a \subseteq \mathfrak{H}_a \subseteq \mathfrak{G}_a$, *then the group* $\mathfrak{H}$ *generated by the groups* $\mathfrak{H}_a$ *in* $\mathfrak{G}$ *is*

7.01. $$\mathfrak{H} = \{\overset{*}{\prod_a} \mathfrak{H}_a ; \mathfrak{U}_{a\beta} = \mathfrak{U}_{\beta a}\}.$$

Further, if, for every α, $\bar{\mathfrak{H}}_a$ *is the group generated by* $\mathfrak{H}$ *and* $\mathfrak{G}_a$ *in* $\mathfrak{G}$, *then* $\mathfrak{G}$ *is the free product of the groups* $\bar{\mathfrak{H}}_a$ *with the amalgamated subgroup* $\mathfrak{H}$,

7.02. $$\mathfrak{G} = \{\overset{*}{\prod_a} \bar{\mathfrak{H}}_a ; \mathfrak{H} = \mathfrak{H}\}.$$

Proof. For the proof of 7.01 we note first, that $\mathfrak{H}$ contains $\mathfrak{U}$; and as the groups $\mathfrak{U}_a$ have the same significance for the groups $\mathfrak{H}_a$ as for the $\mathfrak{G}_a$, the first part of 5.7 is satisfied by $\mathfrak{H}$.

Next we have to choose suitably the normal form 5.1 for the elements of $\mathfrak{H}$. We choose it as the normal form in $\mathfrak{G}$:

In every group $\mathfrak{G}_a$, we denote by $H_\lambda{}^a$ those representatives of $\mathfrak{G}_a$ with respect to $\mathfrak{U}_a$, which belong to $\mathfrak{H}_a$; for the others we keep the letter $F_\lambda{}^a$. As $\mathfrak{U}_a$ belongs to $\mathfrak{H}_a$, the elements $H_\lambda{}^a$ form exactly a full system of right-hand representatives for $\mathfrak{H}_a$ with respect to $\mathfrak{U}_a$.

An element 5.1 of $\mathfrak{G}$ belongs to $\mathfrak{G}_a$ if, and only if, it is of one of the forms U_aH^a or U_aF^a;[14] it belongs to $\mathfrak{H}_a$ if, and only if, it is of the form U_aH^a. As $\mathfrak{H}$ is generated by the groups $\mathfrak{H}_a$, the normal form in $\mathfrak{G}$ for an element of $\mathfrak{H}$ is such that only representatives H occur in it. Conversely, every normal form of this kind represents an element of $\mathfrak{H}$, because $\mathfrak{U}$ belongs to $\mathfrak{H}$. I.e., $\mathfrak{H}$ consists exactly of those elements of $\mathfrak{G}$ whose normal form 5.1 contains representatives H only. Therefore, different representations of this kind represent different elements of $\mathfrak{H}$. But as this representation also constitutes a normal form *in* $\mathfrak{H}$ for the elements of $\mathfrak{H}$, the second condition of 5.7 is also satisfied. Which proves 7.01.

$\bar{\mathfrak{H}}_a$, which is generated by $\mathfrak{H}$ and $\mathfrak{G}_a$, consists exactly of all those elements of $\mathfrak{G}$ whose normal form 5.1 contains, apart from representatives H, only

[14] To save suffixes, we shall in general omit the suffixes λ which distinguish between different representatives of the same group $\mathfrak{G}_a$.

representatives F^{α} out of this one group $\mathfrak{G}_{\alpha}$. Since the representation 5.1 is unique, any element common to $\bar{\mathfrak{H}}_{\alpha}$ and $\bar{\mathfrak{H}}_{\beta}$ has a normal form which does not contain any representatives $F^{\alpha} \neq 1$ or $F^{\beta} \neq 1$, i.e. it belongs to $\mathfrak{H}$. Since, on the other hand, $\mathfrak{H}$ belongs to every group $\bar{\mathfrak{H}}_{\alpha}$, we have

$$\bar{\mathfrak{H}}_{\alpha} \cap \bar{\mathfrak{H}}_{\beta} = \mathfrak{H} \text{ for every pair } \alpha, \beta.$$

In order to show that $\mathfrak{G}$ is the free product of the groups $\bar{\mathfrak{H}}_{\alpha}$ with the amalgamated subgroup $\mathfrak{H}$, we apply again 5.7. As $\mathfrak{G}$ is generated by the groups $\bar{\mathfrak{H}}_{\alpha}$, and any two of them have the same meet $\mathfrak{H}$ in $\mathfrak{G}$, we need only show that if the elements of $\mathfrak{G}$ are represented in normal form, with respect to the subgroups $\bar{\mathfrak{H}}_{\alpha}$ and $\mathfrak{H}$, then different normal forms represent different elements of $\mathfrak{G}$.

Now the normal form 5.1, in the case of groups with one amalgamated subgroup, reduces to

$$R = UL_1L_2 \cdots L_r.$$ [15]

Hence, we have to choose in every group $\bar{\mathfrak{H}}_{\alpha}$ a system of right-hand representatives with respect to $\mathfrak{H}$; we denote its elements by $\bar{G}^{\alpha}$. With these, every element of $\mathfrak{G}$ can be represented in the form

7.03. $$G = \bar{H}\bar{G}^{\alpha_1}\bar{G}^{\alpha_2} \cdots \bar{G}^{\alpha_n}, \qquad n \geq 0,$$

with $\bar{H}$ in $\mathfrak{H}$, $\bar{G}^{\alpha_\nu} \neq 1$ for $\nu = 1, \cdots, n$, and $\alpha_\nu \neq \alpha_{\nu+1}$. And we have to show that formally different representations 7.03 represent different elements of $\mathfrak{G}$.

To this end, we choose the representatives $\bar{G}^{\alpha}$ in the following way:

First we choose in every group $\mathfrak{G}_{\alpha}$ a system of right-hand representatives $\bar{F}_0{}^{\alpha} = 1$, $\bar{F}^{\alpha}$ for $\mathfrak{G}_{\alpha}$ with respect to $\mathfrak{H}_{\alpha}$. If, as before, the right-hand representatives for $\mathfrak{H}_{\alpha}$ with respect to $\mathfrak{U}_{\alpha}$ are denoted by $H_0{}^{\alpha} = 1$, H^{α}, then all the different products $H^{\alpha}\bar{F}^{\alpha}$ form exactly a full system of right-hand representatives for $\mathfrak{G}_{\alpha}$ with respect to $\mathfrak{U}_{\alpha}$. With these representatives, the normal form 5.1 for the elements of $\mathfrak{G}$ (i.e., as elements of the generalized free product of the groups $\mathfrak{G}_{\alpha}$ with amalgamated subgroups $\mathfrak{U}_{\alpha\beta}$) becomes:

7.04. $$G = U_0(H^{\alpha_1}\bar{F}^{\alpha_1})U_1 \cdots (H^{\alpha_m}\bar{F}^{\alpha_m})U_m, \qquad m \geq 0,$$

where for every $\mu = 1, \cdots, m$ at least one of the two representatives H^{α_μ} and $\bar{F}^{\alpha_\mu}$ is not the unit element. And here we know, that two representations 7.04 represent the same element of $\mathfrak{G}$ if, and only if, they are identical.

G belongs to $\bar{\mathfrak{H}}_{\alpha}$ if, and only if, in 7.04 $\bar{F}^{\alpha_\mu} = 1$ whenever $\alpha_\mu \neq \alpha$.

[15] This is exactly the normal form used by Schreier in [6].

i. e., only representatives $\bar{F}^{\alpha}$ occur. Out of these elements, we can choose as representatives $\bar{G}^{\alpha}$ for $\bar{\mathfrak{H}}_{\alpha}$ with respect to $\mathfrak{H}$ all those formally different elements whose representation 7.04 has the property that $U_0 = H^{\alpha_1} = 1$ (hence $\bar{F}^{\alpha_1} \neq 1$) and $\alpha_1 = \alpha$, i. e. all the elements

7.05. $$\bar{G}^{\alpha} = \bar{F}^{\alpha} U_1 (H^{\alpha_2} \bar{F}^{\alpha_2}) U_2 \cdot \cdot \cdot (H^{\alpha_m} \bar{F}^{\alpha_m}) U_m,$$

where $\bar{F}^{\alpha} \neq 1$, and $\bar{F}^{\alpha_\mu} = 1$ whenever $\alpha_\mu \neq \alpha$.

This is easily seen: On the one hand, every element of $\bar{\mathfrak{H}}_{\alpha}$ is the product of an element of $\mathfrak{H}$—*viz.* the longest part in the beginning of the representation 7.04, which belongs to $\mathfrak{H}$—and an element 7.05. On the other hand, two formally different elements, $\bar{G}_1{}^{\alpha}$ and $\bar{G}_2{}^{\alpha}$, of the form 7.05 belong to different classes with respect to $\mathfrak{H}$. If they did not, we would have:

$$\bar{G}_1{}^{\alpha} = \bar{H} \bar{G}_2{}^{\alpha} \qquad (\bar{H} \text{ in } \mathfrak{H});$$

but if here $\bar{H}$ is written in the form 7.04, then the formal product $\bar{H} \bar{G}_2{}^{\alpha}$ is itself a normal form 7.04, hence equal to that of $\bar{G}_1{}^{\alpha}$ only if $\bar{H} = 1$, $\bar{G}_1{}^{\alpha} \equiv \bar{G}_2{}^{\alpha}$.

Now we read the representation 7.04 of the elements G of $\mathfrak{G}$ in the following way: Beginning from the left, we split off the longest part which belongs to $\mathfrak{H}$. The rest will begin with a representative $\bar{F}^{\alpha} \neq 1$. Then starting with this, we split off the longest part product which belongs to $\bar{\mathfrak{H}}_{\alpha}$; this is exactly a representative $\bar{G}^{\alpha} \neq 1$ of the form 7.05. The remaining part will now begin with a representative $\bar{F}^{\beta} \neq 1$, with $\beta \neq \alpha$. So we go on. Then G will appear represented in the form 7.03, where the representatives $\bar{G}^{\alpha_\nu}$ are of the form 7.05.

On the other hand, every formal product 7.03 leads, by virtue of 7.05, to a representation 7.04; and two representations 7.03 lead to the same representation 7.04 only if they are identical. Hence, two representations 7.03 represent the same element of $\mathfrak{G}$ if, and only if, they are identical. Which proves 7.02.

In the case where the groups $\mathfrak{H}_{\alpha}$ coincide with the $\mathfrak{U}_{\alpha}$, the groups $\bar{\mathfrak{H}}_{\alpha}$ are the subgroups of $\mathfrak{G}$ generated by $\mathfrak{U}$ and $\mathfrak{G}_{\alpha}$; $\mathfrak{H}$ is the subgroup $\mathfrak{U}$ of $\mathfrak{G}$. Hence:

7.06. Corollary. *The generalized free product $\mathfrak{G}$ of the groups $\mathfrak{G}_{\alpha}$ is the free product of its subgroups $\bar{\mathfrak{H}}_{\alpha}$, each generated by $\mathfrak{U}$ and $\mathfrak{G}_{\alpha}$, with the one amalgamated subgroup $\mathfrak{U}$.*

With the help of 7.01, applied to a free product of a finite number of factors $\mathfrak{G}_{\iota}$ with one amalgamated subgroup $\mathfrak{U}$, we can replace the criterion 6.21 by the following:

7.1. CRITERION. *If* $\mathfrak{G}$ *is generated by the groups* $\mathfrak{G}_1, \cdots, \mathfrak{G}_n$ *such that any two of them have the same meet* $\mathfrak{U}$ *in* $\mathfrak{G}$, *then* $\mathfrak{G}$ *is the free product of the groups* $\mathfrak{G}_1, \cdots, \mathfrak{G}_n$ *with the one amalgamated subgroup* $\mathfrak{U}$ *if, and only if,*

$$\mathfrak{G} = \mathfrak{G}^i \underset{\mathfrak{U}}{*} \mathfrak{G}_i \text{ for } i = 1, \cdots, n$$

where, for every i, $\mathfrak{G}^i$ *is the subgroup of* $\mathfrak{G}$ *which is generated by all the groups* $\mathfrak{G}_k$ $(k \neq i)$.

Proof. The condition is necessary by 6.02. That it is sufficient we prove by induction.

The sufficiency is trivial for $n = 2$. Let it be true for $n-1$ factors $\mathfrak{G}_i$; then we have to prove it for n factors $\mathfrak{G}_i$.

For $i = 1, \cdots, n-1$, we denote by $\mathfrak{G}^{i,n}$ the group generated by the $n-2$ groups $\mathfrak{G}_k$ $(k \neq i, n)$ in $\mathfrak{G}$. $\mathfrak{G}^{i,n}$ is a subgroup of $\mathfrak{G}^i$ which contains $\mathfrak{U}$, and $\mathfrak{G}^{i,n}$ and $\mathfrak{G}_i$ together generate $\mathfrak{G}^n$. From

$$\mathfrak{G} = \mathfrak{G}^i \underset{\mathfrak{U}}{*} \mathfrak{G}_i \text{ for } i = 1, \cdots, n-1$$

follows therefore by 7.01:

$$\mathfrak{G}^n = \mathfrak{G}^{i,n} \underset{\mathfrak{U}}{*} \mathfrak{G}_i \text{ for } i = 1, \cdots, n-1.$$

Hence by induction:

$$\mathfrak{G}^n = \{ \overset{*}{\prod_{i=1,\ldots,n-1}} \mathfrak{G}_i ; \mathfrak{U} = \mathfrak{U} \}$$

Since also

$$\mathfrak{G} = \mathfrak{G}^n \underset{\mathfrak{U}}{*} \mathfrak{G}_n,$$

the criterion 6.21 gives

$$\mathfrak{G} = \{ \overset{*}{\prod_{i=1\ldots n}} \mathfrak{G}_i ; \mathfrak{U} = \mathfrak{U} \}.$$

8. While in the generalized free product $\mathfrak{G}$ the group generated by subgroups of the factors $\mathfrak{G}_a$ which contain $\mathfrak{U}_a$, is itself the generalized free product of these subgroups, this is no longer generally true if the subgroups of the $\mathfrak{G}_a$ do not contain the groups $\mathfrak{U}_a$. This paragraph gives some results and examples which show what can happen in this case. Their meaning will become clearer from the general results of Part II, for which they also serve as useful illustrations. We bring them here, as they fall naturally into the context of the preceding paragraph; also, we shall need one of the results, 8.11, in **9**.

To begin with, we restrict ourselves to free products of a finite number of factors with one amalgamated subgroup, and, in order to avoid unessential complications, even to products of two factors only.

8.0. THEOREM. *Let $\mathfrak{G}$ be the free product of $\mathfrak{G}_1$ and $\mathfrak{G}_2$ with the amalgamated subgroup $\mathfrak{U}$. For $i = 1, 2$, let $\mathfrak{H}_i$ be subgroups of $\mathfrak{G}_i$ whose meet with $\mathfrak{U}$ is $\mathfrak{V}_i$. Denote by $\mathfrak{V}_{12}$ the meet of $\mathfrak{H}_1$ and $\mathfrak{H}_2$, i. e.*

$$\mathfrak{H}_1 \cap \mathfrak{H}_2 = \mathfrak{V}_1 \cap \mathfrak{V}_2 = \mathfrak{V}_{12}.$$

Then the group $\mathfrak{H}$ generated by $\mathfrak{H}_1$ and $\mathfrak{H}_2$ in $\mathfrak{G}$ is $\mathfrak{H} = \mathfrak{H}_1 \underset{\mathfrak{V}_{12}}{*} \mathfrak{H}_2$, provided that:

1. *for $i = 1, 2$, the group $\bar{\mathfrak{H}}_i$ generated by $\mathfrak{H}_i$ and $\mathfrak{V}_k$ $(i \neq k)$ in $\mathfrak{G}$ is: $\bar{\mathfrak{H}}_i = \mathfrak{H}_i \underset{\mathfrak{V}_{12}}{*} \mathfrak{V}_k$,*

2. *The meet $\mathfrak{V}$ of $\mathfrak{H}$ and $\mathfrak{U}$ is generated by $\mathfrak{V}_1$ and $\mathfrak{V}_2$, and $\mathfrak{V} = \mathfrak{V}_1 \underset{\mathfrak{V}_{12}}{*} \mathfrak{V}_2$.*

Whether these rather detailed conditions can be satisfied, depends, of course, largely on the nature of the group $\mathfrak{U}$. But in the special case that $\mathfrak{V}_1 = \mathfrak{V}_2$, the meet of $\mathfrak{H}$ with $\mathfrak{U}$ is $\mathfrak{V} = \mathfrak{V}_1 = \mathfrak{V}_2$, and the conditions 1. and 2. are certainly satisfied. Hence:

8.11. COROLLARY. *If $\mathfrak{H}_1$ and $\mathfrak{H}_2$ are subgroups of $\mathfrak{G}_1$ and $\mathfrak{G}_2$ respectively which have the same meet $\mathfrak{V}$ with $\mathfrak{U}$, then $\mathfrak{H}_1$ and $\mathfrak{H}_2$ generate in $\mathfrak{G}$ the free product of $\mathfrak{H}_1$ and $\mathfrak{H}_2$ with the amalgamated subgroup $\mathfrak{V}$.*

Therefore, in particular:

8.12. *If $\mathfrak{H}_1$ and $\mathfrak{H}_2$ are subgroups of $\mathfrak{G}_1$ and $\mathfrak{G}_2$ respectively whose meet with $\mathfrak{U}$ is the unit element, then they generate in $\mathfrak{G}$ the free product of $\mathfrak{H}_1$ and $\mathfrak{H}_2$.*

We now prove Theorem 8.0:

Proof. For $i = 1, 2$, we choose a system $V_0{}^i = 1$, V^i of right-hand representatives for $\mathfrak{V}_i$ with respect to $\mathfrak{V}_{12}$; and a system $H_0{}^i = 1$, H^i of right-hand representatives for $\mathfrak{H}_i$ with respect to $\mathfrak{V}_i$. Then all the different products V^iH^i form a system of right-hand representatives for $\mathfrak{H}_i$ with respect to $\mathfrak{V}_{12}$. Hence all the elements of $\mathfrak{H}$ can be expressed in the form:

8.01. $$H = V_{12}(V^{i_1}H^{i_1}) \cdots (V^{i_r}H^{i_r}), \; r \geq 0, \; V_{12} \text{ in } \mathfrak{V}_{12},$$

with $i_\rho \neq i_{\rho+1}$ for $\rho = 1, \cdots, r-1$; and not both, V^{i_ρ} and H^{i_ρ}, are the unit element, for $\rho = 1, \cdots, r$.

All we have to show is that formally different expressions 8.01 represent different elements of $\mathfrak{H}$.

Now those elements 8.01 where only representatives H^1 of $\mathfrak{H}_1$, or H^2

of $\mathfrak{H}_2$ occur, are exactly all the elements of the free products $\mathfrak{H}_1 \underset{\mathfrak{B}_{12}}{*} \mathfrak{B}_2 = \bar{\mathfrak{H}}_1$ and $\mathfrak{H}_2 \underset{\mathfrak{B}_{12}}{*} \mathfrak{B}_1 = \bar{\mathfrak{H}}_2$ respectively, and therefore in each case different from each other as elements of $\mathfrak{G}$. Moreover, we see in the same way as in the proof of 7.02, that all $\bar{H}^i$ of $\bar{\mathfrak{H}}_i$ whose representation 8.01 begins with a representative $H^i \neq 1$, form a full system of right-hand representatives for $\bar{\mathfrak{H}}_i$ with respect to its subgroup $\mathfrak{B} = \mathfrak{B}_1 \underset{\mathfrak{B}_{12}}{*} \mathfrak{B}_2$. But since $\bar{\mathfrak{H}}_i$ belongs to $\mathfrak{G}_i$, it follows from the condition 2. that also any two representatives $\bar{H}^1 \neq 1$ and $\bar{H}^2 \neq 1$ are different elements of $\mathfrak{G}$.

This shows that, if again as in the proof of 7.02, we read 8.01 in the form

8.02. $$H = V\bar{H}^{i_1}\bar{H}^{i_2}\cdots\bar{H}^{i_s}, \quad s \geq 0, \quad V \text{ in } \mathfrak{B},$$

where $i_\sigma \neq i_{\sigma+1}$ for $\sigma = 1, \cdots, s-1$, then two expressions 8.02 are identical if, and only if, the corresponding expressions 8.01 are identical.

Again from condition 2. we see that both groups $\bar{\mathfrak{H}}_1$ and $\bar{\mathfrak{H}}_2$ have meet $\mathfrak{B}$ with $\mathfrak{U}$. Since they are subgroups of $\mathfrak{G}_1$ and $\mathfrak{G}_2$ respectively, the representatives $\bar{H}^i$ for their classes with respect to $\mathfrak{B}$ are, as elements of $\mathfrak{G}_i$, also in different classes with respect to $\mathfrak{U}$. They can, therefore, be chosen as part of the full system of representatives for $\mathfrak{G}_i$ with respect to $\mathfrak{U}$. But then all the expressions 8.02 are normal forms in $\mathfrak{G} = \mathfrak{G}_1 \underset{\mathfrak{U}}{*} \mathfrak{G}_2$, and therefore represent the same element of $\mathfrak{G}$ if, and only if, they are identical. As, moreover, two expressions 8.02 are identical if, and only if, the corresponding expressions 8.01 are identical, these latter represent the same element of $\mathfrak{G}$, i.e. of $\mathfrak{H}$, if, and only if, they are identical. Which proves 8.0.

We add some remarks on the conditions 1. and 2. of 8.0:

If $\mathfrak{H}$ is the free product of $\mathfrak{H}_1$ and $\mathfrak{H}_2$ with the amalgamated subgroup $\mathfrak{B}_{12}$, then by 7.01, also the subgroups $\mathfrak{H}_i$ and $\mathfrak{B}_k$ of $\mathfrak{H}_i$ and $\mathfrak{H}_k$ respectively (each of which contains $\mathfrak{B}_{12}$) generate their free product with the amalgamated subgroup $\mathfrak{B}_{12}$. Therefore, the condition 1. is certainly necessary, and for the same reason also that part of the condition 2. which requires that $\mathfrak{B}_1$ and $\mathfrak{B}_2$ form in $\mathfrak{U}$ their free product with the amalgamated subgroup $\mathfrak{B}_{12}$—this is, in fact, implied by 1., again because of 7.01. But the full condition 2., that the meet of $\mathfrak{H}$ and $\mathfrak{U}$ is exactly this group $\mathfrak{B}$, is by no means necessary, as the following example shows:

8.21. *Example.* Let $\mathfrak{G} = \{a, b, c, d\}$,[16] and consider the subgroups $\mathfrak{G}_1$ and $\mathfrak{G}_2$ of $\mathfrak{G}$, where

[16] We use the notation used by Magnus [3], i.e., write $\mathfrak{G} = \{a_1, a_2, \cdots; r_k(a_i) = 1\}$, if $\mathfrak{G}$ is the group generated by $a_1, a_2, \cdots$ with the defining relations $r_k(a_i) = 1$.

$$\mathfrak{G}_1 = \{a, b, c\} \text{ and } \mathfrak{G}_2 = \{a, b, (bc)^2, d\}.$$

Then

$$\mathfrak{G} = \mathfrak{G}_1 \underset{\mathfrak{U}}{*} \mathfrak{G}_2$$

with

$$\mathfrak{U} = \{a, b, (bc)^2\}.$$

As subgroups $\mathfrak{H}_1$ and $\mathfrak{H}_2$ of $\mathfrak{G}_1$ and $\mathfrak{G}_2$ respectively we take

$$\mathfrak{H}_1 = \{a, c\} \text{ and } \mathfrak{H}_2 = \{b, d\}.$$

Hence

$$\mathfrak{B}_1 = \{a\},\ \mathfrak{B}_2 = \{b\}, \text{ and } \mathfrak{B}_{12} = 1,$$

so that $\mathfrak{B}_1$ and $\mathfrak{B}_2$ generate their free product $\mathfrak{B}_1 * \mathfrak{B}_2 = \{a, b\}$ in $\mathfrak{U}$. But the group $\mathfrak{H}$ generated by $\mathfrak{H}_1$ and $\mathfrak{H}_2$ is the whole group $\mathfrak{G}$, so that the meet of $\mathfrak{H}$ and $\mathfrak{U}$ is the whole group $\mathfrak{U}$ which is larger than $\mathfrak{B}$. All the same $\mathfrak{H}$ is the free product of $\mathfrak{H}_1$ and $\mathfrak{H}_2$.

Although the condition 2. is not necessary, the condition 1. (from which the necessary part of 2. follows, as we saw above) is not sufficient:

8. 22. *Example.* Let

$$\mathfrak{G} = \{a, b, c, d;\ (abc)^2 = (abd)^2\},$$ [16]

and

$$\mathfrak{G}_1 = \{a, b, c\}, \qquad \mathfrak{G}_2 = \{a, b, d\}.$$

Then

$$\mathfrak{G} = \mathfrak{G}_1 \underset{\mathfrak{U}}{*} \mathfrak{G}_2,$$

where

$$\mathfrak{U} = \{a, b, e\} \text{ with } e = (abc)^2 = (abd)^2.$$

Further let

$$\mathfrak{H}_1 = \{a, c\} \text{ and } \mathfrak{H}_2 = \{b, d\},$$

hence

$$\mathfrak{B}_1 = \{a\},\ \mathfrak{B}_2 = \{b\}, \text{ and } \mathfrak{B}_{12} = 1.$$

The condition 1. is satisfied, for

$$\bar{\mathfrak{H}}_1 = \{a, b, c\} = \mathfrak{H}_1 * \mathfrak{B}_2 = \mathfrak{G}_1,$$

and

$$\bar{\mathfrak{H}}_2 = \{a, b, d\} = \mathfrak{H}_2 * \mathfrak{B}_1 = \mathfrak{G}_2;$$

but the group $\mathfrak{H}$ generated by $\mathfrak{H}_1$ and $\mathfrak{H}_2$ is the whole group $\mathfrak{G}$ which is not the free product of $\mathfrak{H}_1$ and $\mathfrak{H}_2$.

One might expect the condition 2. alone to be sufficient, in which case it would have to follow from the fact that the meet of $\mathfrak{H}$ and $\mathfrak{U}$ is generated by $\mathfrak{B}_1$ and $\mathfrak{B}_2$ and is the free product of $\mathfrak{B}_1$ and $\mathfrak{B}_2$ with amalgamated $\mathfrak{B}_{12}$, that also $\mathfrak{H}_1$ and $\mathfrak{B}_2$, and $\mathfrak{H}_2$ and $\mathfrak{B}_1$, form in $\mathfrak{G}$ their free product with amalgamated $\mathfrak{B}_{12}$. But this is not true either, as is shown by the following example:

8. 23. *Example.* Let

$$\mathfrak{G} = \{a, b, c, d;\ (abc)^2 = 1\},$$

and

$$\mathfrak{G}_1 = \{a, b, c;\ (abc)^2 = 1\} \text{ and } \mathfrak{G}_2 = \{a, b, d\}.$$

Then

$$\mathfrak{G} = \mathfrak{G}_1 \underset{\mathfrak{U}}{*} \mathfrak{G}_2$$

with

$$\mathfrak{U} = \{a, b\}.$$

Further take

$$\mathfrak{H}_1 = \{a, c\} \text{ and } \mathfrak{H}_2 = \{b, d\};$$

then

$$\mathfrak{B}_1 = \{a\},\ \mathfrak{B}_2 = \{b\}, \text{ and } \mathfrak{B}_{12} = 1.$$

Hence

$$\mathfrak{B} = \{a, b\} = \mathfrak{B}_1 * \mathfrak{B}_2 = \mathfrak{U}$$

is the meet of $\mathfrak{H}$ and $\mathfrak{U}$. But again $\mathfrak{H}$ is the whole group $\mathfrak{G}$ which is not the free product of $\mathfrak{H}_1$ and $\mathfrak{H}_2$.

We return once more to the Corollary 8. 11. This expresses a property of any free product with one amalgamated subgroup independently of the nature of this subgroup. We are going to prove it once again, for any number of factors $\mathfrak{G}_\alpha$, in order to show how it follows very easily from the existence of the normal form 5. 1 for the elements of $\mathfrak{G}$.

If again, for every α, the representatives of $\mathfrak{G}_\alpha$ with respect to $\mathfrak{U}$ are denoted by G^α, then all the expressions

$$G = UG^{\alpha_1}G^{\alpha_2}\cdots G^{\alpha_r} \qquad (U \text{ in } \mathfrak{U},\ G^{\alpha_\rho} \neq 1,\ \alpha_\rho \neq \alpha_{\rho+1})$$

represent uniquely all the elements of $\mathfrak{G}$.

The subgroups $\mathfrak{H}_\alpha$ of $\mathfrak{G}_\alpha$ are such, that for every α

$$\mathfrak{H}_\alpha \cap \mathfrak{U} = \mathfrak{B}.$$

Let H^α denote representatives of $\mathfrak{H}_\alpha$ with respect to $\mathfrak{B}$. Then every element of $\mathfrak{H}_\alpha$ is of the form VH^α, hence every element of $\mathfrak{H}$ can be written in the form

$$H = VH^{\alpha_1}H^{\alpha_2}\cdots H^{\alpha_s} \qquad (V \text{ in } \mathfrak{H};\ H^{\alpha_\sigma} \neq 1;\ \alpha_\sigma \neq \alpha_{\sigma+1});$$

and every product of this form is an element of $\mathfrak{H}$.

But two representatives $H_1{}^\alpha \neq 1$ and $H_2{}^\alpha \neq 1$ are also in different classes of $\mathfrak{G}_\alpha$ with respect to $\mathfrak{U}$, as $\mathfrak{H}_\alpha \cap \mathfrak{U} = \mathfrak{V}$. For every α, these H^α can, therefore, be chosen as part of the system of representatives for $\mathfrak{G}_\alpha$ with respect to $\mathfrak{U}$; but then, the above representation of the elements of $\mathfrak{H}$ is the normal form of these elements in $\mathfrak{G}$. Hence, different expressions of this form represent different elements of $\mathfrak{G}$, i. e. of $\mathfrak{H}$; so that, by 5. 7, $\mathfrak{H}$ is the free product of the groups $\mathfrak{H}_\alpha$ with the amalgamated subgroup $\mathfrak{V}$.

From this proof one might expect that a corresponding result holds in the generalized free product $\mathfrak{G} = \{\overset{*}{\prod_\alpha} \mathfrak{G}_\alpha; \mathfrak{U}_{\alpha\beta} = \mathfrak{U}_{\beta\alpha}\}$. If the groups $\mathfrak{H}_\alpha$ are subgroups of $\mathfrak{G}_\alpha$ so that, for every α, β,

$$\mathfrak{H}_\alpha \cap \mathfrak{U}_{\alpha\beta} = \mathfrak{H}_\beta \cap \mathfrak{U}_{\alpha\beta} = \mathfrak{V}_{\alpha\beta},$$

and therefore

$$\mathfrak{H}_\alpha \cap \mathfrak{H}_\beta = \mathfrak{V}_{\alpha\beta},$$

do the groups $\mathfrak{H}_\alpha$ generate in $\mathfrak{G}$ their generalized free product with amalgamated subgroups $\mathfrak{V}_{\alpha\beta}$?

The answer is, in general, negative. The reason that the criterion 5. 7 is useless in this case, is that it is useless in the, loosely speaking, smallest generalized free product $\mathfrak{U}$: The normal form 5. 1 has a meaning only if the groups $\mathfrak{G}_\alpha$ are not all equal to $\mathfrak{U}_\alpha$, it in no way reflects the fact that also $\mathfrak{U}$ is a generalized free product. A similar representation of the elements of $\mathfrak{U}$ cannot be given; and it is, in fact, in $\mathfrak{U}$ that the result in question need not be true, i. e., if the groups $\mathfrak{H}_\alpha$ are subgroups of $\mathfrak{U}_\alpha$ (see the example below). But, if the groups $\mathfrak{H}_\alpha$ are such that their common parts with the groups $\mathfrak{U}_\alpha$ behave reasonably in $\mathfrak{U}$, then so do the whole groups $\mathfrak{H}_\alpha$ in $\mathfrak{G}$. More precisely, it can be shown:

If the meet of $\mathfrak{H}_\alpha$ *and* $\mathfrak{U}_\alpha$ *is* $\mathfrak{W}_\alpha$ *(for every* α*), so that* $\mathfrak{W}_\alpha \cap \mathfrak{U}_{\alpha\beta} = \mathfrak{W}_\beta \cap \mathfrak{U}_{\alpha\beta} = \mathfrak{V}_{\alpha\beta}$, *hence* $\mathfrak{H}_\alpha \cap \mathfrak{H}_\beta = \mathfrak{W}_\alpha \cap \mathfrak{W}_\beta = \mathfrak{V}_{\alpha\beta}$, then the groups $\mathfrak{H}_\alpha$ *generate in* $\mathfrak{G}$ *the free product* $\mathfrak{H}$ *of the groups* $\mathfrak{H}_\alpha$ *with amalgamated subgroups* $\mathfrak{V}_{\alpha\beta}$ *provided that:*

1. *The group* $\mathfrak{W}$ *generated by the groups* $\mathfrak{W}_\alpha$ *in* $\mathfrak{U}$ *is the generalized free product of the* $\mathfrak{W}_\alpha$ *with amalgamated* $\mathfrak{V}_{\alpha\beta}$;

2. *the meet of* $\mathfrak{W}$ *with* $\mathfrak{U}_\alpha$ *is exactly* $\mathfrak{W}_\alpha$, *for every* α.

We mention this fact without proof; it is of no great interest beyond the indication it gives that if difficulties arise in $\mathfrak{G}$, then they arise in the generalized free product $\mathfrak{U}$ contained in $\mathfrak{G}$. That in $\mathfrak{U}$ difficulties do, in fact, arise, has been mentioned already. The following two examples show, that neither of the conditions above need be satisfied; in particular, the generalization of 8.11 need not be true in $\mathfrak{U}$.

8.31. *Example.* Let

$$\mathfrak{U} = \{a, b, c\},$$

and

$$\mathfrak{U}_1 = \{a, b\}; \ \mathfrak{U}_2 = \{a, c\}; \ \mathfrak{U}_3 = \{b, c\}.$$

Then

$$\mathfrak{U} = \{ \prod_i^* \mathfrak{U}_i ; \mathfrak{U}_{ik} = \mathfrak{U}_{ki} \}$$

with

$$\mathfrak{U}_{12} = \mathfrak{U}_{21} = \{a\}; \ \mathfrak{U}_{13} = \mathfrak{U}_{31} = \{b\}; \ \mathfrak{U}_{23} = \mathfrak{U}_{32} = \{c\}.$$

Now we take as subgroups $\mathfrak{W}_i$ of $\mathfrak{U}_i$ the groups

$$\mathfrak{W}_1 = \{a^2, b^2, ab\}, \ \mathfrak{W}_2 = \{a^2, c^2, ac\}, \ \mathfrak{W}_3 = \{b^2, c^2, c^{-1}b\}.$$

For every pair i, k, the groups $\mathfrak{W}_i$ and $\mathfrak{W}_k$ have the same meet $\mathfrak{V}_{ik}$ with $\mathfrak{U}_{ik}$, *viz.*:

$$\mathfrak{V}_{12} = \mathfrak{V}_{21} = \{a^2\}; \ \mathfrak{V}_{13} = \mathfrak{V}_{31} = \{b^2\}; \ \mathfrak{V}_{23} = \mathfrak{V}_{32} = \{c^2\}.$$

The groups $\mathfrak{W}_i$ $(i = 1, 2, 3)$ generate in $\mathfrak{U}$ the group

$$\mathfrak{W}^* = \{a^2, b^2, c^2, d_1, d_2, d_3; \ d_1 = d_2 d_3\} \text{ with } d_1 = ab, \ d_2 = ac, \ d_3 = c^{-1}b;$$

whereas the free product of $\mathfrak{W}_1$, $\mathfrak{W}_2$, and $\mathfrak{W}_3$ with amalgamated subgroups $\mathfrak{V}_{ik}$ is the group

$$\mathfrak{W} = \{a^2, b^2, c^2, d_1, d_2, d_3\}.$$

On the other hand, if the groups $\mathfrak{W}_i$ are such that they generate the generalized free product $\mathfrak{W}$ in $\mathfrak{U}$, the meet of $\mathfrak{W}$ with one or more of the factors $\mathfrak{U}_i$ of $\mathfrak{U}$ need not be the original subgroup $\mathfrak{W}_i$ of $\mathfrak{U}_i$, but may be larger:

8.32. *Example.* We consider the same groups $\mathfrak{U}$, $\mathfrak{U}_1$, $\mathfrak{U}_2$, $\mathfrak{U}_3$ as in the preceding example; we also choose $\mathfrak{W}_1$ and $\mathfrak{W}_2$ as before,

$$\mathfrak{W}_1 = \{a^2, b^2, ab\}; \ \mathfrak{W}_2 = \{a^2, c^2, ac\};$$

but as the third group we choose $\mathfrak{W}^*_3 = \{b^2, c^2\}$ in $\mathfrak{U}_3$. These three groups

generate the same group $\mathfrak{W}^*$ in $\mathfrak{U}$, as the three groups $\mathfrak{W}_1$, $\mathfrak{W}_2$, $\mathfrak{W}_3$. But now $\mathfrak{W}^* = \{a^2, b^2, c^2, d_1, d_2\}$ is the free product of $\mathfrak{W}_1$, $\mathfrak{W}_2$, $\mathfrak{W}^*_3$ with amalgamated subgroups $\mathfrak{V}_{ik}$.

Here the meet of $\mathfrak{W}^*$ with $\mathfrak{U}_3$ is $\mathfrak{W}_3$, which is larger than $\mathfrak{W}^*_3$.

9. We now return to the problem mentioned in **4**: Certain groups $\mathfrak{U}_i$ with their subgroups $\mathfrak{U}_{ik}$ are given; under what condition does their generalized free product exist?

We assume, of course, that the necessary conditions 3. 0 are satisfied. Because of Theorem 5. 0, we may also assume that the groups $\mathfrak{U}_i$ are generated by their subgroups $\mathfrak{U}_{ik}$ (i fixed). In the simplest case of only three factors we can establish two sufficient conditions for the existence of $\mathfrak{U}$.

Let us recall the notation:

Every group $\mathfrak{U}_i$ is generated by its subgroups $\mathfrak{U}_{ik}$ and $\mathfrak{U}_{ij}$ ($i = 1, 2, 3$; $k \neq i \neq j$); $\mathfrak{U}_{ik}$ and $\mathfrak{U}_{ki}$ are isomorphic, and $\mathfrak{U}_{ik} = \mathfrak{J}_{ki}(\mathfrak{U}_{ki})$. The meet of $\mathfrak{U}_{ij}$ and $\mathfrak{U}_{ik}$ in $\mathfrak{U}_i$ is mapped by the isomorphisms $\mathfrak{J}_{ij}$ and $\mathfrak{J}_{ik}$ onto the meet of $\mathfrak{U}_{ji}$ and $\mathfrak{U}_{jk}$ in $\mathfrak{U}_j$, and the meet of $\mathfrak{U}_{ki}$ and $\mathfrak{U}_{kj}$ in $\mathfrak{U}_k$ respectively. We denote this meet by the same letter $\mathfrak{U}_{123}$ in all three groups $\mathfrak{U}_i$.

Then we proceed to prove:

9. 0. THEOREM. $\mathfrak{U} = \{\overset{*}{\prod} \mathfrak{U}_i; \mathfrak{U}_{ik} = \mathfrak{U}_{ki}\}$ *exists whenever one of the groups* $\mathfrak{U}_i$, $\mathfrak{U}_3$ *say, is the free product of its subgroups* $\mathfrak{U}_{31}$ *and* $\mathfrak{U}_{32}$ *with the amalgamated subgroup* $\mathfrak{U}_{123}$.

9. 1. THEOREM. $\mathfrak{U} = \{\overset{*}{\prod} \mathfrak{U}_i; \mathfrak{U}_{ik} = \mathfrak{U}_{ki}\}$ *exists whenever two of the groups* $\mathfrak{U}_i$, $\mathfrak{U}_1$, *and* $\mathfrak{U}_2$ *say, have the property that every element of* $\mathfrak{U}_{ik}$ *is permutable with every element of* $\mathfrak{U}_{ij}$ ($i = 1, 2$).

Proof. We prove these two facts by means of the Criterion 6. 20. This criterion, applied to a generalized free product $\mathfrak{U}$, becomes slightly simpler: As $\mathfrak{G}_i$ coincides with $\mathfrak{U}_i$, which is contained in $\mathfrak{G}^i$, the free product of $\mathfrak{G}^i$ and $\mathfrak{G}_i$ with the amalgamated subgroup $\mathfrak{U}_i$ is simply the group $\mathfrak{G}^i$, and this is a certain factor group of the generalized free product $\mathfrak{F}^i$ of all the factors with the exception of $\mathfrak{G}_i$.

In our case we choose $\mathfrak{G}_i$ as the group $\mathfrak{U}_3$. Then we have to form first

$$\mathfrak{F}^3 = \mathfrak{U}_1 \underset{\mathfrak{U}_{12}}{*} \mathfrak{U}_2.$$

$\mathfrak{F}^3$ contains the subgroups $\mathfrak{U}_{13}$ and $\mathfrak{U}_{23}$ of $\mathfrak{U}_1$ and $\mathfrak{U}_2$ respectively, and because of

$$\mathfrak{U}_{13} \cap \mathfrak{U}_{12} = \mathfrak{U}_{23} \cap \mathfrak{U}_{12} = \mathfrak{U}_{123},$$

these groups $\mathfrak{U}_{13}$ and $\mathfrak{U}_{23}$ generate in $\mathfrak{F}^3$, by 8.11, the group

$$\bar{\mathfrak{U}}^3 = \mathfrak{U}_{13} \underset{\mathfrak{U}_{123}}{*} \mathfrak{U}_{23}.$$

Because of the identifying relations

$$u_{13} = u_{31} = \mathfrak{J}_{13}(u_{13}) \text{ and } u_{23} = u_{32} = \mathfrak{J}_{23}(u_{23})$$

for all elements u_{13} and u_{23} of $\mathfrak{U}_{13}$ and $\mathfrak{U}_{23}$ respectively, $\mathfrak{U}$, if it exists, is obtained from $\mathfrak{F}^3$ by introducing the defining relations of $\mathfrak{U}_3$, expressed in the generators u_{13} and u_{23} instead of the generators u_{31} and u_{32} of $\mathfrak{U}_3$.

Since $\mathfrak{U}_3$ is generated by its subgroups $\mathfrak{U}_{31}$ and $\mathfrak{U}_{32}$ whose meet in $\mathfrak{U}_3$ is $\mathfrak{U}_{123}$, these relations and their conjugates generate in $\bar{\mathfrak{U}}^3$ a self-conjugate subgroup $\mathfrak{r}$ which is tidy with respect to $\mathfrak{U}_{13}$ and $\mathfrak{U}_{23}$ (*cf.* 2). If we denote by $\mathfrak{R}$ the smallest self-conjugate subgroup of $\mathfrak{F}^3$ which contains $\mathfrak{r}$, we have to form the factor group $\mathfrak{F}^3/\mathfrak{R}$. This will be isomorphic to the free product of the groups $\mathfrak{U}_i$ with the amalgamated subgroups $\mathfrak{U}_{ik}$, provided that

1. $\mathfrak{R} \cap \bar{\mathfrak{U}}^3 = \mathfrak{r}$;

2. $\mathfrak{R}$ is tidy with respect to $\mathfrak{U}_1$ and $\mathfrak{U}_2$; and it does not contain any elements of the form $U_i\bar{U}_3$ where U_i is an element of $\mathfrak{U}_i$, but not of $\mathfrak{U}^3$ (i. e. not of $\mathfrak{U}_{i3}$), for $i = 1, 2$, and $\bar{U}_3$ in $\bar{\mathfrak{U}}^3$.

For condition 1. expresses the fact that in $\mathfrak{F}^3/\mathfrak{R}$ the groups $\mathfrak{U}_{13}$ and $\mathfrak{U}_{23}$ generate a group isomorphic to $\mathfrak{U}_3$. And 2. sees to it, that $\mathfrak{U}_1$ and $\mathfrak{U}_2$ are isomorphically represented in $\mathfrak{F}^3/\mathfrak{R}$, such that no two groups $\mathfrak{U}_i$ and $\mathfrak{U}_k$ have a meet larger than $\mathfrak{U}_{ik}$.

In the case of 9.0, we have $\mathfrak{r} = 1$, hence $\mathfrak{R} = 1$, and 1. and 2. are trivially satisfied.

In the case of 9.1, we know that

$$u_{12}u_{13} = u_{13}u_{12}$$

and

$$u_{12}u_{23} = u_{23}u_{12}$$

for all elements u_{12}, u_{13}, and u_{23} of the groups $\mathfrak{U}_{12}$, $\mathfrak{U}_{13}$, and $\mathfrak{U}_{23}$ respectively. Since every element r of $\mathfrak{r}$ is a word in the generators u_{13} and u_{23}, we have, therefore,

$$u_{12}^{-1}ru_{12} = r$$

for every element u_{12} of $\mathfrak{U}_{12}$. Moreover, since $\mathfrak{r}$ is self-conjugate in $\bar{\mathfrak{U}}^3$, we have for $i = 1, 2$ and any element u_{i3} of $\mathfrak{U}_{i3}$:

$$u_{i3}^{-1}ru_{i3} = r' \text{ with } r' \text{ in } \mathfrak{r}.$$

Since every element of $\mathfrak{F}^3$ is a word in the generators u_{12}, u_{13}, and u_{23}, it follows:

For every element U of $\mathfrak{F}^3$, we have

$$U^{-1}rU = r' \text{ with } r' \text{ in } \mathfrak{r}$$

But $\mathfrak{R}$ is generated by all the elements r of $\mathfrak{r}$ and their conjugates in $\mathfrak{F}^3$; therefore

$$\mathfrak{R} = \mathfrak{r}.$$

Hence the condition 1. is certainly satisfied. But 2. also is: $\mathfrak{r}$ is tidy with respect to $\mathfrak{U}_{13}$ and $\mathfrak{U}_{23}$ in $\bar{\mathfrak{U}}^3$, and therefore as subgroup of $\mathfrak{F}^3$, *a fortiori* tidy with respect to $\mathfrak{U}_1$ and $\mathfrak{U}_2$. And if $\mathfrak{r}$ contains an element r of the form

$$r = U_i\bar{U}_3,\ U_i \text{ in } \mathfrak{U}_i,\ \bar{U}_3 \text{ in } \bar{\mathfrak{U}}^3,$$

it follows from the fact that r itself is an element of $\bar{\mathfrak{U}}^3$, that also U_i belongs to $\bar{\mathfrak{U}}^3$, i. e. to $\mathfrak{U}_{i3}$.

This completes the proof of 9. 0 and 9. 1.

One might expect that a suitable generalization of 9. 0 and 9. 1 might hold for any finite number of groups $\mathfrak{U}_i$; e. g. that the generalized free product of n groups $\mathfrak{U}_i$ with amalgamated subgroups $\mathfrak{U}_{ik}$ exists, whenever $n-1$ of them have the property that their subgroups $\mathfrak{U}_{ik}$ (i fixed) are permutable element by element; or, whenever one of the groups, $\mathfrak{U}_n$ say, is itself the free product of its subgroups $\mathfrak{U}_{ni}$ ($i = 1, \cdots, n-1$) with amalgamated subgroups $\mathfrak{U}_{nik} = \mathfrak{U}_{ni} \cap \mathfrak{U}_{nk}$.

That nothing of the kind is true, is shown by the following example of four groups, three of which are free Abelian groups (denoted by their generators in round brackets), and one a free group.

9. 2. *Example.* We consider the following groups $\mathfrak{U}_i$:

$$\mathfrak{U}_1 = (a_{12}, a_{13}) \text{ with } \mathfrak{U}_{12} = \{a_{12}\},\ \mathfrak{U}_{13} = \{a_{13}\},\ \mathfrak{U}_{14} = \{a_{12}a_{13}\}$$

$$\mathfrak{U}_2 = (a_{21}, a_{23}) \text{ with } \mathfrak{U}_{21} = \{a_{21}\},\ \mathfrak{U}_{23} = \{a_{23}\},\ \mathfrak{U}_{24} = \{a_{21}a_{23}\}$$

$$\mathfrak{U}_3 = (a_{31}, a_{32}) \text{ with } \mathfrak{U}_{31} = \{a_{31}\},\ \mathfrak{U}_{32} = \{a_{32}\},\ \mathfrak{U}_{34} = \{a_{31}a_{32}\}$$

$$\mathfrak{U}_4 = \{b_1, b_2, b_3\} \text{ with } \mathfrak{U}_{4i} = \{b_i\} \text{ for } i = 1, 2, 3.$$

These groups satisfy the necessary conditions 3. 0, as in every group $\mathfrak{U}_i$ the meet of any two subgroups $\mathfrak{U}_{ik}$ and $\mathfrak{U}_{ij}$ is the unit element. By 6. 20, the generalized free product of $\mathfrak{U}_1$, $\mathfrak{U}_2$, $\mathfrak{U}_3$, and $\mathfrak{U}_4$ is a homomorphic image of the generalized free product of $\mathfrak{U}_1$, $\mathfrak{U}_2$, and $\mathfrak{U}_3$. The latter exists, by 9. 1, and is obviously

$$\mathfrak{F}^4 = (a_1, a_2, a_3)$$

with

$$a_1 = a_{23} = a_{32},\ a_2 = a_{31} = a_{13},\ a_3 = a_{12} = a_{21}.$$

Hence in $\mathfrak{F}^4$:

$$\mathfrak{U}_{14} = \{a_3a_2\},\ \mathfrak{U}_{24} = \{a_1a_3\},\ \mathfrak{U}_{34} = \{a_2a_1\}.$$

But the subgroup of $\mathfrak{F}^4$ which is generated by these three groups $\mathfrak{U}_{i4}$, is Abelian, so that no homomorphic image of it can be isomorphic with the free group $\mathfrak{U}_4$.

10. We conclude this part by determining the generalized free product of three infinite cycles.

10. 0. THEOREM. *The free product* $\mathfrak{U}$ *of three infinite cycles* $\mathfrak{U}_i$ *with amalgamated subcycles* $\mathfrak{U}_{ik}$ *is itself an infinite cycle.*

Proof. That this generalized free product exists, if only the subgroups $\mathfrak{U}_{ik}$ are chosen so as to satisfy 3. 0, we know by 9. 1. Let

$$\mathfrak{U}_1 = \{a_1\},\ \mathfrak{U}_2 = \{a_2\},\ \mathfrak{U}_3 = \{a_3\}.$$

The subgroups $\mathfrak{U}_{ik}$ are generated by certain powers of a_1, a_2, and a_3 respectively. We put

$$\mathfrak{U}_{12} = \{a_1^{e_{12}}\} \text{ and } \mathfrak{U}_{13} = \{a_1^{e_{13}}\} \text{ in } \mathfrak{U}_1,$$
$$\mathfrak{U}_{23} = \{a_2^{e_{23}}\} \text{ and } \mathfrak{U}_{21} = \{a_2^{e_{21}}\} \text{ in } \mathfrak{U}_2,$$
$$\mathfrak{U}_{31} = \{a_3^{e_{31}}\} \text{ and } \mathfrak{U}_{32} = \{a_3^{e_{32}}\} \text{ in } \mathfrak{U}_3.$$

Since $\mathfrak{U}_{ik}$ and $\mathfrak{U}_{ij}$ together generate $\mathfrak{U}_i$ $(i = 1, 2, 3)$, we have for the greatest common divisors: $(e_{12}, e_{13}) = (e_{23}, e_{21}) = (e_{31}, e_{32}) = 1$. Therefore, the meet of $\mathfrak{U}_{ij}$ and $\mathfrak{U}_{ik}$ in $\mathfrak{U}_i$ is generated by:

$$a_1^{e_{12}e_{13}},\ a_2^{e_{23}e_{21}},\ a_3^{e_{31}e_{32}} \text{ respectively.}$$

To satisfy 3. 0, the identifying relations

$$a_1^{e_{12}} = a_2^{e_{21}},\ a_2^{e_{23}} = a_3^{e_{32}},\ a_3^{e_{31}} = a_1^{e_{13}}$$

must imply:

$$a_1^{e_{12}e_{13}} = a_2^{e_{23}e_{21}} = a_3^{e_{31}e_{32}}.$$

It follows that:

$$(a_1^{e_{12}})^{e_{13}} = (a_2^{e_{21}})^{e_{13}} = (a_2^{e_{21}})^{e_{23}}, \text{ hence: } e_{13} = e_{23} = t;$$
$$(a_1^{e_{13}})^{e_{12}} = (a_3^{e_{31}})^{e_{12}} = (a_3^{e_{31}})^{e_{32}}, \text{ hence: } e_{12} = e_{32} = s;$$
$$(a_2^{e_{23}})^{e_{21}} = (a_3^{e_{32}})^{e_{21}} = (a_3^{e_{32}})^{e_{31}}, \text{ hence: } e_{31} = e_{21} = r.$$

Hence we obtain as the only possible choice of subgroups $\mathfrak{U}_{ik}$:

10.01. $$\mathfrak{U}_{12} = \{a_1^s\} \text{ and } \mathfrak{U}_{13} = \{a_1^t\} \text{ in } \mathfrak{U}_1,$$
$$\mathfrak{U}_{23} = \{a_2^t\} \text{ and } \mathfrak{U}_{21} = \{a_2^r\} \text{ in } \mathfrak{U}_2,$$
$$\mathfrak{U}_{31} = \{a_3^r\} \text{ and } \mathfrak{U}_{32} = \{a_3^s\} \text{ in } \mathfrak{U}_3,$$

10.02. $$\text{with } (r, s) = (s, t) = (t, r) = 1.$$

Now $\mathfrak{U}$ is generated by a_1, a_2, and a_3, with the defining relations:

10.03. $$a_1^s = a_2^r = u_3;\ a_2^t = a_3^s = u_1;\ a_3^r = a_1^t = u_2.$$

Because of 10.02, the generators a_i can be expressed by the u_i, so that also the latter generate $\mathfrak{U}$. But, by 10.03, any two of these, u_i and u_j, are powers of the same generator a_k, and therefore permutable with each other. Hence, $\mathfrak{U}$ is Abelian.

The matrix of exponents belonging to the generators and their defining relations 10.03 is:

$$\begin{pmatrix} s & -r & 0 \\ 0 & t & -s \\ -t & 0 & r \end{pmatrix}.$$

Its determinant is zero, and its elementary divisors are all equal to one, i.e., $\mathfrak{U}$ is an infinite cycle.

It seems likely that also the free product of more than three infinite cycles (but a finite number of them) exists, if only the systems of subcycles satisfy 3.0. But I have not been able to confirm this conjecture.

UNIVERSITY COLLEGE,
HULL, ENGLAND.

REFERENCES.

[1] Brandt, H., *Mathematische Annalen*, vol. 96 (1927), pp. 360-366.
[2] Kurosch, A., *Mathematische Annalen*, vol. 109 (1934), pp. 647-660.
[3] Magnus, W., *Mathematische Annalen*, vol. 105 (1931), pp. 52-74.
[4] Malcev, A., *Mathematische Annalen*, vol. 113 (1937), pp. 686-691.
[5] Reidemeister, K., *Einführung in die kombinatorische Topologie*, Braunschweig, 1932.
[6] Schreier, O., *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, vol. 5 (1927), pp. 161-183.

LATTICES OF CONTINUOUS FUNCTIONS II.*

By Irving Kaplansky.

1. Introduction. Let X be a compact Hausdorff space and $C(X)$ the set of real continuous functions on X. In [3] it was shown that, as a lattice alone, $C(X)$ determines X up to a homeomorphism. In this paper we present two investigations suggested by this result. (1) The lattice automorphisms of $C(X)$ are studied. A complete description (Theorem 1) is given for those lattice automorphisms which are bicontinuous in the topology of uniform convergence; an example shows that discontinuous lattice automorphisms also exist. (2) It is natural to seek to characterize those lattices which are of the form $C(X)$ for some X. This we have not done, but as a step in that direction we give such a characterization for "translation lattices," i. e., lattices with a supplementary operation which corresponds to the addition of constant functions.

These investigations were of course suggested by similar ones that have been carried out for rings, Banach spaces, and lattice-ordered groups. It is perhaps worth remarking that the various devices used there (ideals, functionals, and l-ideals, respectively) cease to be available in the present context, and the full burden is thrown upon the natural lattice-theoretic tool: the (lattice) prime ideal.

2. Lattice automorphisms. By a prime ideal in a lattice is meant the inverse image of 0 in a lattice homomorphism on the two-element lattice $(0, I)$. If X is a compact Hausdorff space, then [3, Lemma 3] any prime ideal P in $C = C(X)$ is associated with a unique point x of X, in the sense that $f \varepsilon P$ and $g(x) < f(x)$ imply $g \varepsilon P$. Now let σ be a lattice automorphism of C. It follows from [3, Lemmas 4 and 5] that σ carries prime ideals associated with the same point, say x, into prime ideals again associated with a single point which we shall call $\theta(x)$. Thus σ induces a well defined mapping $x \to \theta(x)$ which is one-to-one and covers all of X. By [3, Lemma 7], θ is a homeomorphism of X.

To describe more closely the relation between σ and θ we make a further restrictive assumption. We topologize C by uniform convergence; to be explicit, we may use the norm $\| f \| = \sup | f(x) |$. We shall suppose that σ is bicontinuous in this topology. As a preliminary, we prove two lemmas.

* Received September 30, 1947.

LEMMA 1. *Let C be topologized by uniform convergence. Then a closed prime ideal P in C has the following form: for some point x in X and some real number α, P consists of all f with $f(x) \leqq \alpha$. Conversely every set of this form is a closed prime ideal.*

Proof. For any prime ideal P, whether closed or not, we can assert the following: for a certain point x and real number α, all functions f with $f(x) < \alpha$ are in P and all with $f(x) > \alpha$ are not in P. Now a function satisfying $f(x) = \alpha$ can be exhibited as a uniform limit of functions f_i with $f_i(x) < \alpha$. Hence if P is closed it contains all f with $f(x) = \alpha$, and so is of the form described. The final statement of the lemma is evident.

LEMMA 2. *Let X be a normal space, and $\psi(x, \alpha)$ a real-valued function defined for real α and $x \varepsilon X$, with the property that $\psi[x, f(x)]$ is a continuous function of x for every continuous real function $f(x)$ and that ψ is continuous in α for fixed x. Then ψ is continuous jointly in x and α.*

Proof. Suppose that ψ is not continuous at (y, β). Then there exists an ϵ such that in every neighborhood of (y, β) we can find a point where ψ differs from $c = \psi(y, \beta)$ by more than 2ϵ. There exists a neighborhood U of β such that $| \psi(y, \alpha) - c | < \epsilon$ for α in U, and, by taking $f(x)$ to be the constant function β, we can find a neighborhood V of y such that $| \psi(x, \beta) - c | < \epsilon$ for x in V. Now take within U a sequence of neighborhoods shrinking to β. In each we may find an α_i, and an x_i in V to go with it, such that $| \psi(x_i, \alpha_i) - c | > 2\epsilon$; moreover it can be arranged that the x's are distinct, since each may be chosen within a neighborhood excluding its predecessors. We define $f(x_i) = \alpha_i$, $f(x) = \beta$ elsewhere in the closure of $\{x_i\}$, and we may extend f to be continuous on all of X since X is normal. Then $\psi[x, f(x)]$ is by hypothesis continuous. If z is a limit point of $\{x_i\}$, then $\psi[z, f(z)] = \psi(z, \beta)$ differs from c by at least 2ϵ. This contradicts the fact that z lies in the closure of V.

We now prove the main theorem.

THEOREM 1. *Let X be a compact Hausdorff space, C the set of real continuous functions on X, and let C be topologized by uniform convergence. Let σ be a lattice automorphism and homeomorphism of C. Then σ has the form $\sigma(f) = f^*$, where $f^*[\theta(x)] = \psi[x, f(x)]$, θ is a homeomorphism of X, and $\psi(x, \alpha)$ is a real-valued function continuous jointly in x and α for $x \varepsilon X$ and α real, and strictly monotone in α for fixed x.*

Proof. The manner in which the homeomorphism θ arises has already

been described. We construct the function ψ as follows. For given x and α, form P, the closed prime ideal consisting of all f with $f(x) \leqq \alpha$. The image $\sigma(P)$ is a closed prime ideal associated with $\theta(x)$; suppose $\sigma(P)$ consists of all g with $g[\theta(x)] \leqq \beta$. Then β is a well defined function of x and α and we write $\beta = \psi(x, \alpha)$. Since σ is a homeomorphism, elements on the frontier of P go into elements on the frontier of $\sigma(P)$. Thus a function f with $f(x) = \alpha$ goes into a function f^* with $f^*[\theta(x)] = \psi(x, \alpha)$. That ψ is jointly continuous in its arguments is a consequence of Lemma 2. If $\alpha > \alpha_0$, then for the corresponding prime ideals P, P_0 we have that P properly contains P_0. Likewise $\sigma(P)$ properly contains $\sigma(P_0)$, and this proves that $\psi(x, \alpha)$ is, for fixed x, strictly monotone in α.

It is true, conversely, that with θ, ψ, and f^* as above, the mapping $f \to f^*$ is a lattice automorphism and homeomorphism of C. The fact that the mapping is one-to-one and order-preserving is clear. The one point needing serious verification is the bicontinuity. Suppose, then, that $f_i \to f$ uniformly. Given ϵ, we may for any fixed x choose i so large that

$$| \psi[x, f_i(x)] - \psi[x, f(x)] | < \epsilon.$$

By the continuity of ψ this will extend to a neighborhood U of x. A finite number of the U's cover X. If we take j to be the maximum of the corresponding i's we have $\| f_j^* - f^* \| < \epsilon$. That the mapping is also continuous in the reverse direction follows from the fact that the transformation $f^* \to f$ has a representation of the same kind, but with different θ and ψ of course.

We may apply Theorem 1 to the more special cases where C is regarded as a ring or lattice-ordered group. It is to be remarked that the topology of uniform convergence is then definable *intrinsically*. We first detect when a function $f > 0$ is strictly positive: in the ring case, if f^{-1} exists; in the lattice-ordered group case, if the multiples of f dominate everything. The neighborhoods of 0 are then sets of the form $-f < a < f$. Thus an automorphism is automatically a homeomorphism. We further observe that in the ring case the only automorphism of the reals is the identity, and in the lattice-ordered group case the only automorphism is multiplication by a positive constant. We deduce the following corollaries.

COROLLARY 1. *All automorphisms of $C(X)$ as a ring are induced by homeomorphisms of X.*

COROLLARY 2. *All automorphisms of $C(X)$ as a lattice-ordered group are induced by homeomorphisms of X followed by multiplication by strictly positive continuous functions.*

The case where $C(X)$ is regarded as a Banach space may also be reduced to the lattice theorem, but there is a preliminary complication, in that a Banach space automorphism is a lattice isomorphism on part of the space and a lattice anti-isomorphism on the remainder. For the complete result, cf. [5, Th. 83].

In the pure lattice case we are unable to define the topology of uniform convergence intrinsically. An equivalent remark is the following: lattice-theoretically we cannot distinguish strictly positive functions from those which vanish on a nowhere dense set. (A function $f > 0$ which vanishes on an open set can be detected through the existence of a second function $g > 0$ such that $f \cap g = 0$).

The following example provides an illustration. Let X be the Stone-Cech compactification of a countably infinite discrete space Y, say for definiteness the positive integers. We define a mapping $f \rightarrow f^*$ on $C(X)$ as follows:

$$f^*(n) = f(n) \text{ if } f(n) \leqq 0,$$

$$f^*(n) = nf(n) \text{ if } 0 \leqq f(n) \leqq n^{-1},$$

$$f^*(n) = f(n) + 1 - n^{-1} \text{ if } f(n) \geqq n^{-1}.$$

It is readily seen that this defines a lattice automorphism of $C(X)$. The function f given by $f(n) = n^{-1}$ vanishes everywhere on $X - Y$, but it is sent into the constant function 1. Moreover the mapping $f \rightarrow f^*$ is not continuous in the topology of uniform convergence, although the inverse mapping is. To get a lattice automorphism which is discontinuous in both directions, take the Cartesian product of X with itself and let the mapping be $f \rightarrow f^*$ on one component and $f^* \rightarrow f$ on the other.

3. Translation lattices. By a translation lattice L we shall mean a lattice where for every $a \varepsilon L$ and real number α a sum $a + \alpha$ is defined. The following postulates are assumed.

1. L is a distributive lattice,
2. $a + 0 = a$,
3. $(a + \alpha) + \beta = a + (\alpha + \beta)$,
4. $\alpha > 0$ implies $a + \alpha > a$,
5. $a > b$ implies $a + \alpha > b + \alpha$.

In addition to these more or less trivial axioms, we assume two deeper ones of an archimedean nature.

6. For any a, $b \varepsilon L$ there exists a real number α such that $a + \alpha > b$,

7. If $a < b + \alpha$ for every positive α, then $a \leqq b$.

By a homomorphism of L into the reals we shall mean a mapping H that satisfies

$$H(a \cup b) = \operatorname{Max}[H(a), H(b)], \tag{1}$$

$$H(a \cap b) = \operatorname{Min}[H(a), H(b)], \tag{2}$$

$$H(a + \alpha) = H(a) + \alpha. \tag{3}$$

Our representation theorem for translation lattices rests primarily on the existence of "sufficiently many" such homomorphisms. In more precise terms we have the following result.

LEMMA 3. *Let L be a translation lattice satisfying axioms 1-7, and c, $d \varepsilon L$ elements such that $c \nleqq d$. Then there exists a homomorphism H of L into the reals such that $H(c) > H(d)$.*

Proof. There must exist a positive α such that $c \nleqq d + \alpha$; otherwise we would have $c \leqq d$ (axiom 7). By [1, Theorem 5.8, or 6, Theorem 6] we may construct a prime ideal P containing $d + \alpha$ but not c. For any a in L we define

$$H(a) = \operatorname*{Inf}_{\beta} (a + \beta \not\varepsilon P). \tag{4}$$

It follows from axiom 6 that for large positive β we have $a + \beta \not\varepsilon P$, and for β negative and large in absolute value, $a + \beta \varepsilon P$. Thus (4) yields a well defined finite real number for $H(a)$. That H satisfies (3) is evident. We next verify (1), the verification of (2) being of course similar. Suppose, for definiteness, that $H(a) \geqq H(b)$ and write $H(a) = \gamma$. For any $\epsilon > 0$, $a - \gamma - \epsilon$ and $b - \gamma - \epsilon$ are in P and hence so is

$$(a - \gamma - \epsilon) \cup (b - \gamma - \epsilon) = (a \cup b) - (\gamma + \epsilon).$$

(We are using here an easy consequence of axioms 2-5). It follows that $H(a \cup b) \leqq \gamma + \epsilon$. Since it is evident that $H(a \cup b) \geqq H(a) = \gamma$, we have $H(a \cup b) = \gamma$, as desired. Finally $H(c) \geqq 0$, $H(d + \alpha) \leqq 0$, so that $H(c) > -\alpha \geqq H(d)$.

We now derive a functional representation for translation lattices. This representation has the peculiarity that, even after suitable topological completion, we do not get all functions but instead a certain subset as described in Theorem 2. (The possibility of getting all functions is of course also

admitted and corresponds to every α being $-\infty$, every $\beta=\infty$.) The inevitability of this kind of restriction is clear from the fact that it is invariant under addition of constant functions and under the lattice operations, and so is automatically compatible with the axioms of a translation lattice.

THEOREM 2. *Let L be a translation lattice satisfying axioms 1-7. Then L is isomorphic, as a translation lattice, to a dense subset under uniform convergence of a set C of functions defined as follows. There is a compact Hausdorff space X; for every pair of distinct points $x, y \in X$ we are given $\alpha(x,y)$ and $\beta(x,y)$ with*

$$-\infty \leqq \alpha(x,y) < \beta(x,y) \leqq \infty;$$

and C is to consist of all real continuous functions f on X satisfying

$$\alpha(x,y) \leqq f(x) - f(y) \leqq \beta(x,y) \tag{5}$$

for all x, y. (The equality is of course excluded if α and/or β is infinite).

Proof. Select arbitrarily an element u in L, and let X denote the set of homomorphisms of L into the reals which map u into 0. In a natural way, the elements of L are represented as real-valued functions on X, and by Lemma 3 the representation is one-to-one. Moreover the representation carries the lattice operations of L into the join and meet of functions, and carries addition of real numbers in L into addition of constant functions.

We topologize X in a well known way. Given H_0 in X, we take an element a in L and a positive number ϵ, and form the set of all H in X satisfying $|H(a) - H_0(a)| < \epsilon$. These sets are defined to constitute a sub-base of the open sets in X. One proves that X is a compact Hausdorff space by showing it to be a closed subset of a Cartesian product of intervals; we omit the familiar details of this argument. Finally the very definition of the topology assures us that the elements of L are represented as continuous functions on X.

It remains to be seen just what functions do arise. Let x, y be any distinct points of X. The element u takes the value 0 at both x and y. Besides this there must exist some element in L taking distinct values at x and y, for x and y are different homomorphisms. By addition of a suitable constant we pass to a function in L vanishing at y but not at x. Next we remark the following: suppose that we have elements g, h in L satisfying $g(y) = h(y) = 0$, $\gamma = g(x) < h(x) = \delta$, and let $\gamma < \zeta < \delta$. Then L contains a function, namely $(g - \gamma + \zeta) \cup h$, which vanishes at y and takes the value ζ at x. Thus the entire interval (γ, δ) is attained at x by functions vanishing at y.

We now let $\alpha(x,y)$, $\beta(x,y) = \operatorname{Inf} f(x)$, $\operatorname{Sup} f(x)$ for all $f \varepsilon L$ vanishing at y. Then it is clear from the preceding remark that for any μ, ν satisfying

$$\alpha(x,y) < \mu - \nu < \beta(x,y), \tag{6}$$

L contains a function f with $f(x) = \mu$, $f(y) = \nu$. Finally we need only quote the theorem of Weierstrass-Stone-Kakutani [cf. 2, Lemma 7.2] to deduce that C, the closure of L under uniform convergence, contains all functions satisfying (5). (It is to be observed that when we pass to the closure, the strict inequality of (6) becomes equality in the event that α and/or β is finite).

4. Further remarks. (a) If we wish to obtain in Theorem 2 the conclusion that we have all functions, unencumbered by a restriction such as (5), we must assume an additional operation which allows us to "stretch" the functions. The assumption of addition or of multiplication by real numbers will do. A weaker assumption than either of these is a doubling operation $a \to 2a$, suitably axiomatized. With some such addendum, Theorem 2 subsumes the representation theorem of Stone [7] for lattice-ordered groups, the operation of addition of reals corresponding to the addition of scalar multiples of Stone's unit element e. (Actually, the existence of these scalar multiples of e is not postulated in Stone's case, but it is an immediate consequence of the assumed completeness. To facilitate comparison of the two theorems, we might have postulated addition only of rational numbers in a translation lattice).

(b) It is to be remarked that in the proof of Theorem 2 only one arbitrary choice had to be made, namely the selection of u. In other words: the only automorphism of $C(X)$ as a translation lattice is a homeomorphism of X followed by the addition of a fixed function.

(c) The restrictive condition (5) resembles closely the condition which arises in Kakutani's representation of (M)-spaces without unit [3, Th. 1]; in Kakutani's case the relation between the values of a function at two points is however multiplicative rather than additive. In fact there is a generalization of Kakutani's theorem analogous to Theorem 2: one assumes a lattice in which there is an additional operation of *multiplication* by real numbers. We shall not attempt to give the details except to remark that the same lattice-theoretic tools are used in the proof.

(d) Some remarks can be made in case the archimedean axioms 6 and 7 do not hold. If a, b are such that $a < b + \alpha$ and $b < a + \alpha$ for all positive α, then a and b cannot be distinguished by homomorphisms into the reals, and

the functional representation necessarily identifies them. This passage to sets of equivalent elements is the analogue of reduction modulo a radical. Again if axiom 7 fails, we may still set up a functional representation if we admit $\pm\infty$ as functional values. (Cf. [7] and [8] for similar remarks on lattice-ordered groups).

(e) We conclude with some observations on the possibility of replacing the reals by some more general object R. Lattice automorphisms of $C(X)$ may be studied for any chain R, but to formulate the notion of uniform convergence used in Theorem 1 we require a concept of uniformity in R. For the proof of Theorem 1 (and more particularly Lemma 2) we moreover require the following axiom: any continuous R-valued function defined on a closed subset of X may be extended to all of X. With these provisos, Theorem 1 remains valid. In the particular case where R is discrete, the hypothesis that σ be bicontinuous is automatically fulfilled and may be deleted; and the extendability property of R-valued functions on X is assured if X is totally disconnected. When R consists of just two elements, this gives us Stone's theorem that any automorphism of a Boolean ring is induced by a homeomorphism of the corresponding Boolean space [5, Th. 4].

To formulate a generalization of Theorem 2 we must take R to be a simply ordered group. But more than that, the proof as given requires that R be complete (or at least completable) with respect to Dedekind cuts; and the only complete simply ordered group is the group of real numbers. Another remark is that the proof of the compactness of the space of homomorphisms uses the local compactness of R.

The Post algebras of Rosenbloom [4] are closely related to translation lattices; in essence we take R to be a cyclic group of order n and modify the axioms slightly. A proof very much like that of Theorem 2 can be given for the following generalization of Stone's [5, Th. 1]: a Post algebra is isomorphic to the set of all continuous functions from a compact totally disconnected space to the set $(0, 1, \cdots, n-1)$. In an oral communication to the author, Rosenbloom has described another proof of this theorem based on a reduction to the case of a Boolean ring.

Added in proof (May 12, 1948). It can be shown that if X satisfies the first axiom of countability, then every lattice automorphism of $C(X)$ is automatically continuous. Certain weaker hypotheses will do, but the example given above shows that some assumption must be made. The proof is briefly as follows. Let f, g be functions which are equal at x. Choose a sequence x_i approaching x, and a function h with $h > f \cup g$ at the even points and

$h < f \cap g$ at the odd points of $\{x_i\}$. If σ is a lattice automorphism of $C(X)$, and θ the associated homeomorphism of X, it can be seen that $\sigma(h)$ coincides with $\sigma(f)$ and $\sigma(g)$ at $\theta(x)$. Thus the value of $\sigma(f)$ at $\theta(x)$ is entirely determined by the value of f at x. It follows that σ has the form described in Theorem 1 and is continuous.

UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY.

1. G. Birkhoff, "Lattice theory," *American Mathematical Society Colloquium Publications*, vol. 25, New York, 1940.

2. S. Kakutani, "Concrete representation of abstract (M)-spaces," *Annals of Mathematics*, vol. 42 (1941), pp. 994-1024.

3. I. Kaplansky, "Lattices of continuous functions," *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 617-623.

4. P. C. Rosenbloom, "Post algebras I," *American Journal of Mathematics*, vol. 64 (1942), pp. 167-188.

5. M. H. Stone, "Applications of the theory of Boolean rings to general topology," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 375-481.

6. M. H. Stone, "Topological representations of distributive lattices and Brouwerian logics," *Casopis Matematiky a Fysiky*, vol. 67 (1937), pp. 1-25.

7. M. H. Stone, "A general theory of spectra II," *Proceedings of the National Academy of Sciences*, vol. 27 (1941), pp. 83-87.

8. K. Yosida, "On the representation of the vector lattice," *Proceedings of the Imperial Academy of Tokyo*, vol. 18 (1942), pp. 339-342.

ON IDEALS OF DIFFERENTIABLE FUNCTIONS.*

By HASSLER WHITNEY.

1. Introduction. We shall consider functions of class C^r defined in an open subset G of Euclidean n-space E^n, that is, functions with continuous partial derivatives through the order r in G. (We take r finite.) A set I of such functions forms an *ideal* if, whenever f_1, f_2 are in I and ϕ_1, ϕ_2, are of class C^r in G, $\phi_1 f_1 + \phi_2 f_2$ is in I. We shall call I an *r-ideal.*

Let I be an r-ideal in G. Given any point $x = (x_1, \cdots, x_n)$ in G, consider all sets of numbers

$$a_{\alpha_1 \ldots \alpha_n} \qquad \text{(each } \alpha_i \geqq 0,\ \alpha_1 + \cdots + \alpha_n \leqq r)$$

such that for each such set there is a function f in I with

$$\frac{\partial^{\alpha_1 + \cdots + \alpha_n} f(x)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} = a_{\alpha_1 \ldots \alpha_n}.$$

These sets form the *local ideal* $\mathfrak{J}_r(I, x)$ *of* I *at* x.

It was conjectured by Laurent Schwartz (personal communication) that an ideal is determined by its set of local ideals, provided that the ideal is closed (we use the topology described below). The main object of this paper is to prove this conjecture (see Theorem I). There is a rather obvious generalization of the theorem to the case where E^n is replaced by a manifold of class C^r.

The case $r = 0$ is much simpler than the case $r > 0$; we treat it separately in **3**.

In the last section we consider briefly the following problem: What sets of local ideals can be the set of local ideals of an r-ideal? A satisfactory answer to this question seems quite difficult to find.

The topology we shall use is determined as follows. Let f be a function of class C^r in G. Then to any compact subset A of G (i. e. bounded closed subset A of E^n lying in G), and any $\epsilon > 0$, the set of all functions g of class C^r in G such that

$$\left| \frac{\partial^{\alpha_1 + \cdots + \alpha_n}(g - f)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} \right| < \epsilon \text{ in } A \qquad (\alpha_1 + \cdots + \alpha_n \leqq r)$$

* Received November 25, 1947.

form a "special neighborhood" of f. A set of functions forms an open set if it can be expressed as the union of a set of special neighborhood of functions.

2. The principal theorems. The main theorem is:

THEOREM 1. *Any two closed r-ideals in G with the same sets of local ideals are identical.*

This is an immediate consequence of the following approximation theorem. For the statement of this theorem, we need some more definitions.

We say that a set of functions forms a *locally finite set of functions in G* if the following are true:

(a) Each function is identically zero outside some compact subset of G.

(b) Each point of G is in a neighborhood such that at most a finite number of the functions fail to vanish identically in this neighborhood.

Because of the Heine-Borel theorem, (b) holds with neighborhoods of points replaced by arbitrary compact subsets of G.

If $f_1, f_2, \cdots$ form a locally finite set, and the f_i are of class C^r in G, clearly $f_1 + f_2 + \cdots$ exists and is of class C^r in G.

If I is an r-ideal in G, we define its *locally finite completion* I^* as follows. A function f is in I^* if there is an infinite series of functions of I which converges to f, these functions forming a locally finite set.

It is easily seen that I^* is an ideal. It is not hard to show that $I \subset I^*$; we do not need this fact.

THEOREM 2. *Let I be an r-ideal in G, and let F be a function of class C^r in G such that at each x in G, the values of F and its derivatives form a set of numbers in $\mathfrak{F}_r(I, x)$. Then for each positive continuous function $\epsilon(x)$ in G there is a function f^* in I^* such that*

$$\left| \frac{\partial^{\alpha_1+\cdots+\alpha_n}[F(x) - f^*(x)]}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} \right| < \epsilon(x), \ x \in G, \ \alpha_1 + \cdots + \alpha_n \leqq r.$$

We prove Theorem 1 with the help of Theorem 2 as follows. Let I and I' be two closed r-ideals in G as stated. To show that they are identical, we show that each is contained in the other; for instance, that $I' \subset I$. Take any function F in I'. To prove $F \in I$ it is sufficient to find a function f of I in an arbitrary special neighborhood U of F, since I is closed. Let U be defined by the compact subset A of G and the number $\epsilon > 0$. Set $\epsilon(x) \equiv \epsilon$, and choose f^* by Theorem 2; say $f^* = g_1 + g_2 + \cdots$. As noted above, at most a finite number of these are $\not\equiv 0$ in A; hence

$$f^* = g_1 + \cdots + g_m \text{ in } A \text{ for some } m.$$

Set $f = g_1 + \cdots + g_m$ in G. Then $f \varepsilon I \cap U$, as required.

Most of the paper is devoted to the proof of Theorem 2. The methods used are similar to those employed in extending differentiable functions.[1]

REMARK. The hypothesis of closure in Theorem 1 is necessary, as shown by the first two examples of **4.**

3. The case $r = 0$. This case is described in full by the following theorem.

THEOREM 3. *A closed ideal I of continuous functions in G is determined by the set of points A where all the functions vanish. This set is closed in G. Any closed subset of G can occur.*

The set of points where one function of I vanishes is closed, since the function is continuous. The set A is the intersection of such sets, and hence is closed.

Let A be any closed subset of G. Let $f(x)$ be the distance from any point x of G to A; this is continuous in G, and vanishes in A only. Let I be the ideal generated by f, i. e. the set of all functions ϕf, ϕ continuous. Clearly A is the set of points where all functions of I vanish.

To prove the first part of the theorem, we prove Theorem 2 for this case. Let, then, I and A be as in Theorem 3, and let F be a continuous function such that $F(x) = 0$ in A. Let $\epsilon(x)$ be any positive continuous function in G.

To each x in $G - A$ there corersponds a function f_x in I such that $f_x(x) \neq 0$. By multiplying by a suitable continuous function, we may replace f_x by a function f'_x in I such that for some spherical neighborhood U_x of x (we take $\bar{U}_x \subset G - A$),

$$f'_x(x') > 0 \text{ in } U_x, \qquad f'_x(x') = 0 \text{ in } G - U_x.$$

It is easy to choose $x^1, x^2, \cdots$ so that the spheres $U_{x^1}, U_{x^2}, \cdots$ cover $G - A$, and each intersects but a finite number of the others. Set

$$g'_i(x) = f'_{x^i}(x) / \sum_j f'_{x^j}(x) \text{ in } U_{x^i}, \; g'_i(x) = 0 \text{ in } G - U_{x^i}.$$

Since the numerator is $\equiv 0$ in $G - U_{x^i}$ and the denominator is > 0 in $G - A$, g'_i is continuous. In fact, we may write $g'_i \doteq \phi_i f'_{x^i}$, where

[1] See H. Whitney, "Analytic extensions of differentiable functions defined in closed sets," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 63-89. We refer to this paper as AE.

$\phi_i = 1/\Sigma f'_{x^j}$ in U_{x^i} and $\phi_i = 0$ outside a larger neighborhood of x^i, showing that g'_i is in I. Note that

$$0 \leqq g'_i(x) \leqq 1, \qquad \sum_i g'_i(x) = 1 \text{ in } G - A.$$

Let U be a neighborhood of A such that

$$| F(x) | < \epsilon(x) \text{ in } U.$$

Let $\lambda_1, \lambda_2, \cdots$ be those numbers such that for any $k = \lambda_j$, U_{x^k} has points in $G - U$. Set

$$g_i(x) = F(x) g'_{\lambda_i}(x) \qquad (i = 1, 2, \cdots).$$

Then these form a locally finite set of functions in G, each lying in I. If $f^* = \Sigma g_i$, clearly $f^* = F(x)$ in $G - U$, while for $x \in U$,

$$| F(x) - f^*(x) | = | F(x) | \, | 1 - \Sigma g'_{\lambda_i}(x) | \leqq | F(x) | < \epsilon(x),$$

completing the proof.

4. Examples. We give first two examples of ideals which are not closed.

Example 1. Let I be the set of all functions of one variable, of class C^r, each of which vanishes except in some bounded set. Let $\phi(x) \equiv 1$; ϕ is not in I. Given any bounded set A, let f be a function $\equiv 1$ in A and $\equiv 0$ outside some neighborhood of A; f is in I. Hence, if $\bar{I}$ is the closed ideal, ϕ is in $\bar{I}$. Any function F of class C^r equals $F\phi$, and hence is in $\bar{I}$.

Example 2. Let I consist of all functions of class C^r which vanish in some neighborhood of the origin. Then $\bar{I}$ contains all functions of class C^r which vanish together with their derivatives of order $\leqq r$, at the origin.

The next three examples serve to illustrate what local ideals may be like.

Example 3. $n = 2$, $r = 2$. Let I consist of all functions of class C^r which vanish on the x-axis; we consider the local ideal $\mathfrak{I}_r(I, O)$ of I at the origin O. For f in I, the only necessary relations between derivatives at the origin are

$$f(O) = 0, \qquad \partial f(O)/\partial x = 0, \qquad \partial^2 f(O)/\partial x^2 = 0.$$

Consider the polynomials (which are in I)

$$P_1(x, y) = y, \qquad P_2(x, y) = xy, \qquad P_3(x, y) = y^2.$$

Any function f in I has the same value and derivatives of order $\leqq 2$ at O

as those of some linear combinations of these polynomials; these polynomials (or rather, their sets of values and derivatives) form a base for $\mathfrak{I}(I, O)$.

Note that the local ideal is generated by $P_1 = y$; for $P_2 = xP_1$, $P_3 = yP_1$, and thus P_1 being in I implies that P_2 and P_3 are also.

A new ideal would be obtained by adding the condition $\partial f/\partial y = 0$ at the origin.

EXAMPLE 4. $n = 2$, $r = 2$. Let I consist of all functions of class C^2 which vanish on both axes. Then for any $x^0 \neq O$ on either axis, $\mathfrak{I}(I, x^0)$ is like the local ideal of Example 3. At the origin, $\mathfrak{I}(I, O)$ consists of the set of values 0 alone, except that $\partial^2 f/\partial x \partial y$ may be $\neq 0$; hence the polynomial xy forms a base for $\mathfrak{I}_r(I, O)$.

If we wish the functions to vanish on the x-axis and the line $y = x$, transforming axes (compare Example 5 below) shows that the local ideal is then generated by $xy - y^2$.

EXAMPLE 5. $n = 2$, $r = 2$. Let I now be the set of functions of class C^2 which vanish on the curve $y = x^2$. If we use the variables x', y' in Example 3, we obtain Example 5 by the transformation $x' = x$, $y' = y - x^2$. Calculating derivatives of $f \varepsilon I$ gives, referring to Example 3,

$$\partial f(O)/\partial x' = \partial f(O)/\partial x = 0, \qquad \partial^2 f(O)/\partial x'^2 = 2\,\partial f(O)/\partial y + \partial^2 f(O)/\partial x^2 = \mathbf{0}.$$

The relations defining $\mathfrak{I}_r(I, O)$ are obtained by setting these expressions and $f(O)$ equal to 0. The polynomials

$$P_1 = y - x^2, \qquad P_2 = xy, \qquad P_3 = y^2$$

form a base for $\mathfrak{I}_r(I, O)$, and the local ideal is generated by P_1. As far as $\mathfrak{I}_r(I, O)$ is concerned, P_2 and P_3 are equivalent to xP_1 and yP_1 respectively (see **6**).

The next two examples illustrate the dependence of a local ideal on neighboring values of the functions.

EXAMPLE 6. $n = 2$, $r = 1$. Let A be a point set defined as follows. It contains the origin, and the closed intervals $1/2^{i+1} \leqq x \leqq 1/2^i$ $(y = 0)$ for $i = 1, 3, 5, \cdots$. The omitted intervals $(i = 2, 4, \cdots)$ are replaced by the circular arcs with the same end points, each arc lying in the half plane $y \geqq 0$ and being of radius 1. Let I consist of all functions of class C^1 which vanish on A.

Any function in I has vanishing partial derivatives at the ends of each interval; hence the same is true at the origin. This is in spite of the fact

that the direction between any two points of A sufficiently near to O is as nearly in the x-direction as we please.

EXAMPLE 7. $n=2$, $r=1$. Let us round off the corners in the point set A of Example 6. Then I contains a function f such that $\partial f(O)/\partial y=1$. If we change the original A sufficiently slightly, $\partial f/\partial y$ cannot satisfy a Lipschitz condition.

The last two examples show how the functions in ideals may lack differentiability properties.

EXAMPLE 8. $n=2$, $r=1$. Let I be generated by

$$f(x,y)=[1+(x^2+y^2)^{1/2}]y;$$

then I consists of all functions ϕf, where ϕ is of class C^1. Since $f=0$ only on the x-axis, and $\partial f/\partial y\neq 0$ there, the local ideals are the same as for the ideal I' generated by $f'(x,y)=y$. Since each of f, f' is obtained from the other by multiplying by a function which is not of class C^1, $I\neq I'$.

EXAMPLE 9. $n=3$, $r=1$. Let I be generated by the two functions

$$f(x,y,z)=y+[1+(x^2+y^2+z^2)^{1/2}]z,$$
$$g(x,y,z)=y^2+z^2.$$

Then the set of points where all functions of I vanish is the x-axis. Differentiating shows that

$$\partial f/\partial z=(1+|x|)(\partial f/\partial y) \text{ on the } x\text{-axis}.$$

Since this is trivially true with f replaced by g, we see at once that this relation holds for all functions of I. It follows that there is no function of class C^2 in I for which $\partial f/\partial y\neq 0$ at the origin. The same is of course true in the closed ideal. (A similar remark applies to the last example.)

5. Notations. We shall commonly use single letters to denote n-fold quantities; for $\alpha=(\alpha_1,\cdots,\alpha_n)$, write

$$[\alpha]=\alpha_1+\cdots+\alpha_n.$$

We use D_α for $\partial^{[\alpha]}/\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}$ etc. Thus the right hand side of the second displayed relation in **6**, written out in full, is

$$\sum_{\alpha_1+\cdots+\alpha_n\leqq r}\frac{a_{\alpha_1\ldots\alpha_n}}{\alpha_1!\cdots\alpha_n!}(x_1-x_1^0)^{\alpha_1}\cdots(x_n-x_n^0)^{\alpha_n}.$$

We shall use ρ_{xy} or $\rho(x,y)$ for the distance between x and y; similarly for $\rho_{xA}=\rho(x,A)$ where A is a point set. We set $\rho(x,A)=\infty$ if A is void.

By $U_\delta(x)$, or $U_\delta(A)$, we denote the set of points within δ of the point x or the point set A.

Let f be a function of class C^r. Its *r-value* $V_r(f,x)$ at x is the set of numbers consisting of its value and partial derivatives of order $\leqq r$ at x. Two functions f, g are r-equivalent at x if they have the same r-value there; we write

$$f \approx_r g \text{ at } x, \quad \text{or,} \quad V_r(f,x) = V_r(g,x).$$

We shall call the *grade* of a function f at x^0 the smallest number i such that some partial derivative of f of order i is $\neq 0$ at x^0, provided there is such an i. The grade of a polynomial at the origin is then the lowest degree of its terms if the polynomial is not $\equiv 0$. The grade of an r-value is that of any corresponding function.

6. Local ideals as linear spaces. Given a point x^0, any set a of numbers $\{a_\alpha\}$, $[\alpha] \leqq r$, considered as associated with x^0, is the r-value of a uniquely defined polynomial P of degree $\leqq r$. P must satisfy the condition

$$V_r(P,x^0) = a, \text{ i. e. } D_\alpha P(x^0) = a_\alpha, \ [\alpha] \leqq r.$$

P is given by the formula

$$P(x) = \sum_{[\alpha]\leqq r} (a_\alpha/\alpha!)\,(x - x^0)^\alpha$$

The addition and multiplication of r-values corresponds to the same for polynomials, provided we drop out terms of degree $> r$ when they arise, i. e. reduce mod polynomials of grade $> r$. We may thus represent the r-value by the corresponding polynomial.

As an illustration of the multiplication, at the end of Example 5, **4**, we noted that

$$x(y - x^2) = xy - x^3 \approx_2 xy, \qquad y(y - x^2) \approx_2 y^2, \text{ at } O.$$

We note in passing that x and $x + xy$ generate the same local ideal at the origin: $\mathfrak{I}_r(x, O) = \mathfrak{I}_r(x + xy, O)$. For if x is in an ideal, so is yx, and hence also $x + xy$. Conversely, if $x + xy$ is in an ideal, so are

$$x + xy, \qquad y(x + xy) = xy + xy^2, \cdots, xy^{r-1} + xy^r;$$

hence so is the alternating sum, which is $x \pm xy^r \approx_r x$.

For each point x^0, consider the classes of functions of C^r, two functions being in the same class if they are r-equivalent at x^0. These add and multiply like r-values; hence they may be used to represent r-values. If we are in a

differentiable manifold, they may be used to define r-values and local ideals.

In the proofs of the theorems of this paper, it is addition and multiplication by constants in local ideals which play the essential role. With these operations, a local ideal forms a finite dimensional linear space. A *base* in a local ideal is then a set of r-values in the local ideal such that any r-value in the local ideal can be expressed uniquely as a linear combination of these (with numerical coefficients). A set of r-values forms a base if and only if they are independent and their number equals the dimension of the ideal (as a linear space).

The set of all r-values forms the *unit* local ideal $\mathfrak{J}_r{}^1(x^0)$ at x^0; the set containing the zero r-value (i. e. $a_\alpha = 0$ for each α) only forms the *zero* local ideal $\mathfrak{J}_r{}^0(x^0)$ at x^0. Any local ideal at x^0 contains the second and is contained in the first. If an ideal contains an r-value a with $a_{0\ldots 0} \neq 0$, it is the unit ideal, clearly.

7. On the structure of local ideals. The results of this section will not be needed in the proof of the theorems of the paper.

Consider a point x^0 and an integer s, $0 \leqq s \leqq r$. The set of all r-values $\{a_\alpha\}$ such that $a_\alpha = 0$ if $[\alpha] \neq s$ forms a subspace of $\mathfrak{J}_r{}^1(x^0)$; its dimension is easily seen to be $\binom{n-1+s}{n-1}$. These subspaces are independent and generate $\mathfrak{J}_r{}^1(x^0)$; the dimension of the whole space is $\binom{n+r}{n}$.

Now take any local r-ideal H at x^0. Let H^s be the subset of H consisting of all r-values in H of grade $\geqq s$. Let $\bar{H}^s$ denote the local s-ideal at x^0, formed from H^s by considering a_α with $[\alpha] \leqq s$ only (i. e. by reducing mod r-values of grade $> s$). Then $\bar{H}^s$ is a linear space; let us call its dimension the s-dimension $\dim_s(H)$ of H.

We can relate bases in the $\bar{H}^s$ to a base in H as follows.

LEMMA 7a. *Let H be a local r-ideal at x^0. For each s, $0 \leqq s \leqq r$, let $P_1{}^s, \cdots, P_{d_s}{}^s$ be r-values in H^s such that, when reduced mod r-values of grade $> s$, they form a base in $\bar{H}^s$. Then the $P_i{}^s$ form a base in H, and $d_s = \dim_s(H)$.*

The proof is straightforward, and may be omitted.

8. The sets A_λ, B_λ. We begin by associating with I certain point sets, as follows. Define:

$$A_\lambda = \text{all } x \text{ such that } \dim \mathfrak{J}_r(I, x) = \lambda,$$

$$B_\lambda = \sum_{\mu \geqq \lambda} A_\mu.$$

LEMMA 8a. *The B_λ are open in G.*

Take $x \varepsilon B_\lambda$; say $x \varepsilon A_\mu$, $\mu \geqq \lambda$. Now $\mathfrak{F}_r(I, x)$ is a linear space of dimension μ. Choose μ r-values in it which form a base. For each of these r-values, choose a function in I with this r-value at x^0. Let $f_1, \cdots, f_\mu$ be these functions. Then $V_r(f_1, x^0), \cdots, V_r(f_\mu, x^0)$ are independent. Since the f_i are of class C^r, $V_r(f_i, x^0)$ is a continuous function of x^0; it follows that $V_r(f_1, x), \cdots, V_r(f_\mu, x)$ are independent, x near x^0. This shows that for some neighborhood U of x^0, $\dim \mathfrak{F}_r(I, x) \geqq \mu$ for $x \varepsilon U$, i. e., $U \subset B_\mu \subset B_\lambda$, proving that B_λ is open.

9. Local functions. Let I and F be as in the theorem. For each $x \varepsilon G$ there is a function f in I such that $V_r(f, x) = V_r(F, x)$; it follows from Taylor's formula with remainder that for each $\epsilon > 0$ there is a $\delta > 0$ such that

$$| D_\alpha[f(y) - F(y)] | \leqq \rho_{xy}{}^{r-[\alpha]}\epsilon, \qquad \rho_{xy} < \delta, \ [\alpha] \leqq r.$$

However, δ may depend on x; no uniformity condition holds in general. But if we restrict ourselves to the individual sets A_λ, we can prove a uniformity condition, as shown by the lemmas below.

LEMMA 9a. *For any $x \varepsilon A_\lambda$ and any $\epsilon > 0$ there is a $\delta > 0$ with the following property. For each y in A_λ within δ of x there is a function in I such that if $[\alpha] \leqq r$,*

$$(9.1) \qquad | D_\alpha[f(z) - F(z)] | \leqq \rho_{yz}{}^{r-[\alpha]}\epsilon, \qquad \rho_{xz} < \delta,$$

$$(9.2) \qquad D_\alpha f(y) = D_\alpha F(y).$$

First choose functions $f_1, \cdots f_\lambda$ in I such that $V_r(f_1, x), \cdots, V_r(f_\lambda, x)$ form a base in $\mathfrak{F}_r(I, x)$. Let U be a neighborhood of x such that the r-values of $f_1, \cdots f_\lambda$ are independent at any $y \varepsilon U$. Now take any $y \varepsilon U \cap A_\lambda$. Then the r-values of $f_1, \cdots f_\lambda$ form a base there; hence there are uniquely defined numbers $u_1(y), \cdots, u_\lambda(y)$ such that

$$V_r[u_1(y)f_1 + \cdots + u_\lambda(y)f_\lambda, y] = V_r(F, y), \qquad y \varepsilon U \cap A_\lambda.$$

If we set

$$(9.3) \qquad g_y(z) = u_1(y)f_1(z) + \cdots + u_\lambda(y)f_\lambda(z),$$

this means that

$$(9.4) \qquad D_\alpha g_y(y) = D_\alpha F(y), \qquad y \varepsilon U \cap A_\lambda, \ [\alpha] \leqq r.$$

Choose $\delta_1 > 0$ and N so that

$$| u_i(y) | \leqq N, \qquad y \varepsilon A_\lambda, \ \rho_{xy} < \delta_1.$$

Set $\epsilon_1 = \epsilon/(1+\lambda N)$. By Taylor's formula with remainder, there is a $\delta_2 \leqq \delta_1$ such that, for each k and α,

$$| D_\alpha f_k(z) - \sum_\beta (D_{\alpha+\beta} f_k(y)/\beta!)(z-y)^\beta | \leqq \rho_{yz}{}^{r-[\alpha]}\epsilon_1, \qquad \rho_{xy}, \rho_{xz} < \delta_2.$$

Since $D_\gamma g_y(z) = \sum_k u_k(y) D_\gamma f_k(z)$, this gives

$$\begin{aligned} | D_\alpha g_y(z) &- \sum_\beta (D_{\alpha+\beta} g_y(y)/\beta!)(z-y)^\beta | \\ &= | \sum_k u_k(y)[D_\alpha f_k(z) - \sum_\beta (D_{\alpha+\beta} f_k(y)/\beta!)(z-y)^\beta] | \\ &\leqq \lambda N \rho_{yz}{}^{r-[\alpha]}\epsilon_1, \qquad \rho_{xy}, \rho_{xz} < \delta_2. \end{aligned}$$

Again by Taylor's formula, there is a $\delta \leqq \delta_2$ such that

$$|D_\alpha F(z) - \sum_\beta (D_{\alpha+\beta} F(y)/\beta!)(z-y)^\beta | \leqq \rho_{yz}{}^{r-[\alpha]}\epsilon_1, \qquad \rho_{xy}, \rho_{xz} < \delta.$$

Using (9.4) with $\alpha+\beta$ in place of α and comparing the last two inequalities gives

$$| D_\alpha g_y(z) - D_\alpha F(z) | \leqq (1+\lambda N)\rho_{yz}{}^{r-[\alpha]}\epsilon_1, \qquad \rho_{xy}, \rho_{xz} < \delta.$$

Now take any y in A_λ within δ of x. Set $f(z) = g_y(z)$. Then f is in I, and (9.1) and (9.2) follow from the last relation and (9.4).

LEMMA 9b. *For each positive continuous function $\epsilon(x)$ in A_λ there is a positive continuous function $\eta(x)$ in A_λ with the following property. For each $x \varepsilon A_\lambda$ there is a function f in I such that*

$$(9.5) \qquad | D_\alpha[f(y) - F(y)] | \leqq \rho_{xy}{}^{r-[\alpha]}\epsilon(x), \qquad \rho_{xy} < \eta(x),\ [\alpha] \leqq r,$$

$$(9.6) \qquad D_\alpha f(x) = D_\alpha F(x), \qquad [\alpha] \leqq r,$$

Note that y need not lie in A_λ.

Define the positive continuous function $\epsilon_1(x)$ in A_λ by

$$\epsilon_1(x) = \underset{y \varepsilon U_\theta(x) \cap A_\lambda}{\text{g. l. b.}} \epsilon(y), \qquad \theta = \rho(x, E - B_\lambda)/2.$$

For any x in A_λ, let $\eta(x)$ be the upper bound of numbers η with the following property. There is a point $x' \varepsilon A_\lambda$ such that, using x' and $\epsilon = \epsilon_1(x')$ in Lemma 9a, there is a δ satisfying the condition of the lemma, and also the conditions

$$(9.7) \qquad U_{2\eta}(x) \subset U_\delta(x'), \qquad \delta < \rho(x', E - B_\lambda)/2.$$

Certainly $\eta(x) > 0$. For if we take $x' = x$, we may apply Lemma 9a to find a number $\eta \leqq \delta/2$. To show that $\eta(x)$ is continuous, it is sufficient to show that for any x_1 and x_2 in A_λ,

$$\eta(x_2) \geqq \eta(x_1) - \rho(x_1, x_2)/2.$$

Take any $\theta > 0$, take $\eta > \eta(x_1) - \theta$, and find an x' and δ corresponding to x_1 and η; then clearly x' and δ will also do for x_2, with η replaced by $\eta - \rho_{x_1x_2}/2$, which is arbitrarily near $\eta(x_1) - \rho_{x_1x_2}/2$.

Now take any $x \varepsilon A_\lambda$. Set $\eta = \eta(x)/2$. Using the definition of $\eta(x)$, find x' and δ correspondingly. By (9.7), x is within δ of x'; using $y = x$, let f be the corresponding function in I. Then (9.2) becomes (9.6), while (9.1) gives

$$| D_a[f(z) - F(z)] | \leqq \rho_{xz}{}^{r-[a]}\epsilon_1(x'), \qquad \rho_{x'z} < \delta.$$

Now take any y within $\eta(x)$ of x. Then $y \varepsilon U_{2\eta}(x) \subset U_\delta(x')$; hence the above inequality holds with $z = y$. Moreover,

$$\rho(x, x') < \delta < \rho(x', E - B_\lambda)/2;$$

therefore $\epsilon_1(x') \leqq \epsilon(x)$, and (9.5) is proved.

10. The functions $\phi_\nu(x)$. In each of the next two sections we shall need a process of interpolation between functions of class C^r. For a given closed subset P of E, the functions will be defined in the open set $E - P$ (or in part of it). We follow essentially AE, §§ 7-10.

We shall use a "standard subdivision" of $E - P$, defined as follows. Divide E into n-cubes, each of side 1; let K_0 be the set of all these cubes whose distances from P are at least $6n^{1/2}$ (if there are any). In general, K_s is formed by subdividing all the cubes into cubes of side $1/2^s$, and choosing all those (if any) which do not lie in any of $K_0, \cdots, K_{s-1}$, and whose distances from P are at least $6n^{1/2}/2^s$.

It is easily seen that the distance from any cube Q of K_s to P is $< 18n^{1/2}/2^s$ if $s \geqq 1$, and that any cube of K_s is separated from any cube of K_{s+2} by at least four cubes of K_{s+1}.

Let $y^1, y^2, \cdots$ denote the vertices of all the cubes of $K_0 + K_1 + \cdots$, arranged in a sequence. For each ν, let x_ν be a point of P such that $\rho(x^\nu, y^\nu) = \rho(y^\nu, P)$. Let J_ν be the cube with center y^ν which is just large enough to contain all cubes Q of $K_0 + K_1 + \cdots$ with y^ν as a vertex.

Let Q_0 be a fixed n-cube. Let $\phi'(x)$ be a function of class C^r in E such that

$$\phi'(x) > 0 \text{ within } Q_0, \qquad \phi'(x) = 0 \text{ outside } Q_0.$$

By translating and contracting, we obtain functions $\phi'_\nu(x)$, positive within J_ν and zero outside J_ν. Setting

$$\phi_\nu(x) = \phi'_\nu(x)/\sum_\mu \phi_\mu(x)$$

gives functions of class C^r in E, such that ϕ_ν is positive within J_ν and is zero outside J_ν, and such that

$$\sum_\nu \phi_\nu(x) = 1, \qquad x \varepsilon E - P. \tag{10.1}$$

If J_ν is of side σ, then any derivative of order t of ϕ'_ν is $1/\sigma^t$ times the same derivative of ϕ' at a corresponding point. The definition of ϕ_ν in terms of the ϕ'_μ depends on the relative arrangements of the cubes J_μ about J_ν; there are but a finite number of arrangements. From these facts, we see easily (compare AE, § 10) that for some number $N \geqq 1$,

$$| D_\alpha \phi_\nu(x) | < 2^{s[\alpha]} N, \qquad x \varepsilon Q,\ Q \text{ in } K_s.$$

We shall replace this by another inequality. Take any $x \varepsilon Q$, Q in K_s, $s \geqq 1$. Since $6n^{1/2}/2^s \leqq \rho(Q,P) < 18n^{1/2}/2^s$, and diam $(Q) = n^{1/2}/2^s$, we have

$$\rho(x,P) \leqq \rho(Q,P) + n^{1/2}/2^s \leqq 7\rho(Q,P)/6 < 21n^{1/2}/2^s.$$

Therefore

$$| D_\alpha \phi_\nu(x) | < (21n^{1/2})^r N \rho_{xP}^{-[\alpha]} \tag{10.2}$$

if $x \varepsilon Q$, Q in K_s, $s \geqq 1$, $[\alpha] \leqq r$.

We note that N depends on n and r only, and thus is independent of the shape of P.

We shall need some further properties of the cubes J_ν. Let x be any point of $E - P$, let x lie in the cube Q of K_s, and let x lie within J_ν; then [2]

$$\rho(x,P)/2 < \rho(y^\nu, P) < 2\rho(x,P), \tag{10.3}$$

$$\operatorname{diam}(J_\nu) \leqq 4n^{1/2}/2^s \leqq 2\rho(x,P)/3. \tag{10.4}$$

To prove these, let Q' be a largest cube of $K_0 + K_1 + \cdots$ with y^ν as vertex; say Q' is in K_t. Then $t \geqq s - 1$, and

$$\operatorname{diam}(Q') = n^{1/2}/2^t \leqq 2n^{1/2}/2^s;$$

from which (using the definition of J_ν) the first part of (10.4) follows. It follows from this that $\rho(x, y^\nu) \leqq 2n^{1/2}/2^s$. But also

$$\rho(x,P) \geqq \rho(Q,P) \geqq 6n^{1/2}/2^s;$$

hence $\rho(x,y^\nu) \leqq \rho(x,P)/3$, and (10.3) and the rest of (10.4) follow.

[2] Correction to AE, (9.1): If y^* is distant δ_* from a given point x^0 of A, we prove $d_\nu < 2\delta_*$ only. Note that $d_\nu = r_\nu$.

We note finally that there is a number c such that at most c sets J_ν contain points of any cube Q of $K_0 + K + \cdots$.

11. The functions in the A_λ. In **9** we showed that we could find functions approximating to F locally, with a uniformity condition holding throughout A_λ. In the present section we find a single function approximating to F throughout A_λ, in fact throughout an open subset R_λ of B_λ which contains A_λ. This function will be obtained by interpolating between the functions defined locally.

When a positive continuous function $\zeta(x)$ is given in A_λ, we shall let R_λ denote the set of all x such that for some $y \in A_\lambda$, $\rho(x, y) < \zeta(y)$; R'_λ will denote the same, using $\zeta(y)/2$. It will be clear, with the ζ to be used, that $R_\lambda \subset B_\lambda$.

LEMMA 11a. *Let A_λ be non-void, and let $\epsilon(x)$ be a positive continuous function in B_λ. Then there is a positive continuous function $\eta(x)$ in A_λ such that for each positive continuous function $\zeta(x) \leqq \eta(x)$ in A_λ there is a function f of class C^r in B_λ, given by* (11.10), *the $f_{x\nu}$ being in I, such that if $[\alpha] \leqq r$, then*

$$(11.1) \qquad | D_\alpha[f(x) - F(x)] | \leqq \epsilon(x), \qquad x \in R_\lambda,$$

$$(11.2) \qquad | D_\alpha[f(x) - F(x)] | \leqq \rho_{xA_\lambda}{}^{r-[\alpha]}\epsilon(x), \qquad x \in R_\lambda - R'_\lambda.$$

In the application in **12**, we shall use $\zeta = \eta$.

Using the numbers N and c introduced in **10**, and recalling that $\rho(x, A) = \infty$ if A is void, set

$$(11.3) \qquad \epsilon_1(x) = \epsilon(x)/\{c[(r+1)!]^n(126n^{1/2})^r N\},$$

$$(11.4) \qquad \epsilon_2(x) = \text{g.l.b.}\, \epsilon_1(y), \qquad \rho(x, y) \leqq \min[\rho(x, E - B_\lambda)/2, 1].$$

Using $\epsilon_2(x)$ in A_λ, Lemma 9b gives a positive continuous function in A_λ, which we shall call $\eta_1(x)$; we take

$$(11.5) \qquad \eta_1(x) \leqq \min[\rho(x, E - B_\lambda)/2, 1].$$

Define $\eta_2(x)$ and $\eta(x)$ in A_λ by

$$(11.6) \qquad \eta_2(x) = \text{g.l.b.}\, \eta_1(y), \qquad \rho(x, y) \leqq \min[\rho(x, E - B_\lambda)/2, 4],$$

$$(11.7) \qquad \eta(x) = \min[\eta_2(x)/6, \rho(x, E - B_\lambda)/16, 1/2].$$

Take any $\zeta(x) \leqq \eta(x)$, and define R_λ, R'_λ correspondingly.

Choose a finite or denumerable set A'_λ of points in A_λ, with the following properties:

(a) A'_λ has no limit points in B_λ.

(b) for each $y \in A_\lambda$ there is a $z \in A'_\lambda$ such that

$$\rho(y, z) < \zeta(y)/2, \qquad \zeta(z) > 3\zeta(y)/4.$$

The proof of existence of A'_λ is quite elementary.

We prove three properties of A'_λ:

(c) For each $x \in R_\lambda$ there is a $z \in A'_\lambda$ such that

$$\rho(x, z) < 2\zeta(z).$$

(d) $$\rho(x, A'_\lambda) < \rho(x, E - B_\lambda)/7, \qquad x \in R_\lambda.$$

(e) $$\rho(x, A'_\lambda) < 2\rho(x, A_\lambda), \qquad x \in R_\lambda - R'_\lambda.$$

To prove (c), take $x \in R_\lambda$, choose $y \in A_\lambda$ so that $\rho(x, y) < \zeta(y)$, and choose z by (b). Then

$$\rho(x, z) \leqq \rho(x, y) + \rho(y, z) < 3\zeta(y)/2 < 2\zeta(z),$$

as required. To prove (d) (which is trivial if $B_\lambda = E$), given $x \in R_\lambda$, find y and z as before. Now

$$\rho(x, y) < \zeta(y) \leqq \eta(y) \leqq \rho(y, E - B_\lambda)/16,$$

hence

$$\begin{aligned}\rho(y, E - B_\lambda) &\leqq \rho(x, y) + \rho(x, E - B_\lambda)\\ &\leqq \rho(y, E - B_\lambda)/16 + \rho(x, E - B_\lambda),\end{aligned}$$

and it follows that

$$\rho(y, E - B_\lambda) \leqq 16\rho(x, E - B_\lambda)/15.$$

Using an inequality above, this gives

$$\begin{aligned}\rho(x, A'_\lambda) &\leqq \rho(x, z) < 3\zeta(y)/2 < 2\eta(y)\\ &\leqq \rho(y, E - B_\lambda)/8 \leqq 2\rho(x, E - B_\lambda)/15,\end{aligned}$$

which proves (d). To prove (e), let y be a point of $E - B_{\lambda+1}$ such that $\rho(x, y) = \rho(x, E - B_{\lambda+1})$. Since $E - B_{\lambda+1} = A_\lambda \cup (E - B_\lambda)$, and $\rho(x, A_\lambda) \leqq \rho(x, A'_\lambda) < \rho(x, E - B_\lambda)$, by (d), it follows that

$$y \in A_\lambda, \qquad \rho(x, y) = \rho(x, A_\lambda).$$

Since x is not in R'_λ, $\rho(x,y) \geqq \zeta(y)/2$. Choose z by (b). Then

$$\rho(x,A'_\lambda) \leqq \rho(x,z) \leqq \rho(x,y) + \rho(y,z)$$
$$< \rho(x,y) + \zeta(y)/2 \leqq 2\rho(x,y) = 2\rho(x,A_\lambda),$$

proving (e).

Take now the standard subdivision of the set $B_\lambda - A'_\lambda$; this set is open, because of (a). To each vertex y^ν corresponds a cube J_ν and a function ϕ_ν, as in **10.** We shall, however, use only certain ones of these; we denote by $y^1, y^2, \cdots$ those vertices of the subdivision such that

$$\rho(y^\nu, A'_\lambda) < \rho(y^\nu, E - B_\lambda),$$

and let x^ν, J_ν and ϕ_ν correspond. Then $x^\nu \in A'_\lambda$.

By the choice of $\eta_1(x)$, for each $z \in A'_\lambda$ there is a function $f_z(x)$ in I such that, for $[\alpha] \leqq r$,

$$(11.8) \qquad | D_\alpha[f_z(x) - F(x)] | \leqq \rho_{zx}{}^{r-[\alpha]}\epsilon_2(z), \qquad \rho_{zx} < \eta_1(z),$$

$$(11.9) \qquad D_\alpha f_z(z) = D_\alpha F(z).$$

We define f in B_λ by the formulas

$$(11.10) \qquad f(x) = \begin{cases} \Sigma\phi_\nu(x) f_{x^\nu}(x), & x \in B_\lambda - A'_\lambda, \\ F(x), & x \in A'_\lambda. \end{cases}$$

Clearly f is of class C^r in $B_\lambda - A'_\lambda$. From (11.1) we could deduce at once that f is of class C^{r-1} in B_λ. It requires but little further proof to show that f is of class C^r in B_λ; the necessary details may be found in AE, § 11. (Near any $z \in A'_\lambda$, use (11.1) with $\epsilon(x)$ replaced by an arbitrary $\epsilon > 0$.)

We turn now to the proof (11.1) and (11.2). Take any x in $R_\lambda - A'_\lambda$. Say x is in the cube Q of K_s, and in the interiors of the cubes $J_1, \cdots, J_\gamma$, to use a simple notation. Then $\gamma \leqq c$ (see **10**).

We begin by showing that (the y^i denoting for the moment all vertices of $K_0 + K_1 + \cdots$)

$$(11.11) \qquad \rho(y^i, A'_\lambda) < \rho(y^i, E - B_\lambda), \qquad i = 1, \cdots, \gamma.$$

Set $P = E - B_\lambda + A'_\lambda$. Because of (d),

$$(11.12) \qquad \rho(x, P) = \rho(x, A'_\lambda).$$

Using (10.4) gives (since y^i is the center of J_i)

$$(11.13) \qquad \rho(x, y^i) \leqq \rho(x, P)/3 = \rho(x, A'_\lambda)/3.$$

Hence

$$\rho(y^i, A'_\lambda) \leqq \rho(x, y^i) + \rho(x, A'_\lambda) \leqq 4\rho(x, A'_\lambda)/3.$$

Moreover, using (11.13) and (d) again,

$$\rho(y^i, E - B_\lambda) \geqq \rho(x, E - B_\lambda) - \rho(x, y^i) \geqq 6\tfrac{2}{3}\rho(x, A'_\lambda).$$

These inequalities give (11.11).

It follows from (11.11) that the y^i actually are properly named; further, that the sum in (11.10) contains all ϕ_ν corresponding to $J_1, \cdots, J_\gamma$. Consequently, (10.1) holds at x. Subtracting $F(x)$ from both sides of (11.10) gives therefore (with the present notation)

$$f(x) - F(x) = \sum_{i=1}^{\gamma} \phi_i(x)[f_{x^i}(x) - F(x)]. \tag{11.14}$$

With the same notations again, (11.11), (10.3) and (11.12) give

$$\rho(y^i, x^i) = \rho(y^i, P) < 2\rho(x, P) = 2\rho(x, A'_\lambda).$$

This, with (11.13), gives

$$\rho(x, x^i) < 3\rho(x, A'_\lambda). \qquad i = 1, \cdots, \gamma. \tag{11.15}$$

Using (c), choose $z \varepsilon A'_\lambda$ so that $\rho(x, z) < 2\zeta(z)$. Now

$$\rho(x, x^i) < 3\rho(x, A'_\lambda) \leqq 3\rho(x, z),$$

and hence, using (11.7)

$$\begin{aligned} \rho(z, x^i) < 4\rho(x, z) < 8\zeta(z) &\leqq 8\eta(z) \\ &\leqq \min[\rho(z, E - B_\lambda)/2, 4]. \end{aligned}$$

Consequently, by (11.6), $\eta_2(z) \leqq \eta_1(x^i)$, from which follows, using (11.7) again,

$$\rho(x, x^i) < 3\rho(x, z) < 6\zeta(z) \leqq \eta_2(z) \leqq \eta_1(x^i).$$

We may therefore apply (11.8), giving

$$| D_a[f_{x^i}(x) - F(x)] | \leqq \rho_{x^i x}{}^{r-[a]}\epsilon_2(x^i).$$

Moreover, using (11.5),

$$\rho(x, x^i) < \eta_1(x^i) \leqq \min[\rho(x^i, E - B_\lambda)/2, 1];$$

hence, by (11.4) $\epsilon_2(x^i) \leqq \epsilon_1(x)$. It follows that

$$| D_a[f_{x^i}(x) - F(x)] | \leqq \rho_{x^i x}{}^{r-[a]}\epsilon_1(x), \qquad i = 1, \cdots, \gamma. \tag{11.16}$$

Differentiating (11.14) gives

$$D_\alpha[f(x)-F(x)]=\sum_{i=1}^{\gamma}\sum_{\beta}\binom{\alpha}{\beta}D_\beta\phi_i(x)D_{\alpha-\beta}[f_{x^i}(x)-F(x)].$$

Applying (10.2), (11.12) and (11.16) gives

$$|D_\alpha[f(x)-F(x)]|\leqq\sum_{i=1}^{\gamma}\sum_{\beta}\binom{\alpha}{\beta}(21n^{1/2})^rN\rho_{xA'\lambda}{}^{-[\beta]}\ \rho_{x^ix}{}^{r-[\alpha-\beta]}\epsilon_1(x).$$

(Using an inequality proved below, $\rho(Q,A'_\lambda)\leqq\rho(x,A'_\lambda)\leqq 1<6n^{1/2}$; hence $s\geqq 1$, and the use of (10.2) was legitimate.) Now $\gamma\leqq c$ (see **10**). Each β_k is one of $0,\cdots,r$; hence there are at most $(r+1)^n$ terms in the second sum. Each binomial coefficient is at most $\alpha_1!\cdots\alpha_n!\leqq(r!)^n$. Consequently, using (11.15),

$$(11.17)\qquad |D_\alpha[f(x)-F(x)]|\leqq c[(r+1)!]^n(63n^{1/2})^rN\rho_{xA'\lambda}{}^{r-[\alpha]}\epsilon_1(x),$$

for $x\in R_\lambda-A'_\lambda$, $[\alpha]\leqq r$.

To prove (11.1), take any $x\in R_\lambda-A'_\lambda$. Choosing z as before and applying (c) and (11.7) gives

$$\rho(x,A'_\lambda)\leqq\rho(x,z)<2\zeta(z)\leqq 2\eta(z)\leqq 1;$$

(11.1) is thus an immediate consequence of (11.17) and (11.3), for such x. For $x\in A'_\lambda$, it follows by continuity. (Actually, $D_\alpha f(x)=D_\alpha F(x)$ there.)

To prove (11.2), take $x\in R_\lambda-R'_\lambda$. Then (11.17), property (e) and (11.3) give (11.2) at once. This completes the proof of the lemma.

12. The function f^*. We are given the r-ideal I in G, the function F of class C^r in G such that $V_r(F,x)\in\mathfrak{J}_r(I,x)$ for each $x\in G$, and the positive continuous function $\epsilon(x)$ in G. Define the sets A_λ etc. as in **8**. Let $\lambda_1>\lambda_2>\cdots>\lambda_h$ be those numbers such that each A_{λ_i} is non-void. Set

$$(12.1)\qquad \rho_{x\lambda}=\rho(x,A_\lambda),\qquad \check\rho_{x\lambda}=\rho(x,E-B_\lambda),$$

$$(12.2)\qquad a=1/\{3c[(r+1)!]^n(21n^{1/2})^rN\},$$

and for each λ, let $\epsilon_\lambda(x)$ be the positive continuous function

$$(12.3)\qquad \epsilon_\lambda(x)=\tfrac{1}{2}\min[1,\check\rho_{x\lambda}{}^r]a^{\lambda+1}\epsilon(x),\qquad x\in B_\lambda.$$

For each i, we apply Lemma 11a to the set B_{λ_i}, with the function $\epsilon_{\lambda_i}(x)$. This gives us a function $\eta_{\lambda_i}(x)$ in A_{λ_i}; set $\zeta_{\lambda_i}(x)=\eta_{\lambda_i}(x)$. There is then a function f_{λ_i} in B_{λ_i}, given by (11.10), such that for $[\alpha]\leqq r$,

$$(12.4)\qquad |D_\alpha[f_{\lambda_i}(x)-F(x)]|\leqq\epsilon_{\lambda_i}(x),\qquad x\in R_{\lambda_i},$$

(12.5) $\quad |D_\alpha[f_{\lambda_i}(x) - F(x)]| \leqq \rho_{x\lambda_i}{}^{r-[\alpha]}\epsilon_{\lambda_i}(x), \qquad x \varepsilon R_{\lambda_i} - R'_{\lambda_i}.$

The following relations are immediate consequences:

(12.6) $\quad |D_\alpha[f_{\lambda_i}(x) - F(x)]| \leqq \frac{1}{2}\epsilon(x), \qquad x \varepsilon R_{\lambda_i},$

(12.7) $\quad |D_\alpha[f_{\lambda_i}(x) - F(x)]| \leqq \frac{1}{2}a^{\lambda_{i+1}}\bar{\rho}_{x\lambda_i}{}^{r-[\alpha]}\epsilon(x), \qquad x \varepsilon R_{\lambda_i},$

(12.8) $\quad |D_\alpha[f_{\lambda_i}(x) - F(x)]| \leqq \frac{1}{2}a^{\lambda_{i+1}}\rho_{x\lambda_i}{}^{r-[\alpha]}\epsilon(x), \qquad x \varepsilon R_{\lambda_i} - R'_{\lambda_i}.$

(Consider separately the cases $\bar{\rho}_{x\lambda_i} \leqq 1$, $\bar{\rho}_{x\lambda_i} > 1$.)

Take the standard subdivision of each $B_{\lambda_i} - A_{\lambda_i} = B_{\lambda_{i+1}}$. We shall use only those cubes J_ν which have points not in R_{λ_i}; call these $J_{\lambda_i 1}, J_{\lambda_i 2}, \cdots$. Let $\phi_{\lambda_i j}$ correspond to $J_{\lambda_i j}$, and set

(12.9) $$\phi^*_{\lambda_i}(x) = \sum_j \phi_{\lambda_i j}(x).$$

We show that

(12.10) $\quad \phi^*_{\lambda_i}(x) = 1, \qquad x \varepsilon B_{\lambda_i} - R_{\lambda_i},$

(12.11) $\quad \phi^*_{\lambda_i}(x) = 0, \qquad x \varepsilon R'_{\lambda_i} - A_{\lambda_i}.$

The first relation is clear; see (10.1). To prove the other, we need merely show that no $J_{\lambda_i j}$ has points in R'_{λ_i}. Suppose that this is false; let it have points $x \varepsilon R'_{\lambda_i}$ and $z \varepsilon B_{\lambda_i} - R_{\lambda_i}$. By definition of R'_{λ_i}, there is a point $y \varepsilon A_{\lambda_i}$ such that $\rho(x, y) < \zeta_{\lambda_i}(y)/2$. Also, following the proof of (d), **11**, and using (11.7),

$$\bar{\rho}_{y\lambda_i} \leqq 16\bar{\rho}_{x\lambda_i}/15,$$

$$\rho_{x\lambda_i} \leqq \rho_{xy} < \zeta_{\lambda_i}(y)/2 < \eta_{\lambda_i}(y) \leqq \bar{\rho}_{y\lambda_i}/16 < \bar{\rho}_{x\lambda_i}.$$

Hence, setting $P = E - B_{\lambda_i} + A_{\lambda_i}$, $\rho(x, P) = \rho_{x\lambda_i}$, and (10.4) gives, if $x \varepsilon J_{\lambda_i j}$,

$$\operatorname{diam}(J_{\lambda_i j}) < \rho(x, P) = \rho_{x\lambda_i} \leqq \rho(x, y) < \zeta_{\lambda_i}(y)/2.$$

It follows that $\rho(z, y) \leqq \operatorname{diam}(J_{\lambda_i j}) + \rho(x, y) < \zeta_{\lambda_i}(y)$, so that $z \varepsilon R_{\lambda_i}$, a contradiction.

We note in passing that the proof above (taking $x \varepsilon R_{\lambda_i}$), with the fact $A_{\lambda_{i+1}} \subset E - B_{\lambda_i}$, shows that

(12.12) $\quad \rho_{x\lambda_i} < \bar{\rho}_{x\lambda_i} \leqq \rho_{x\lambda_{i+1}}, \qquad x \varepsilon R_{\lambda_i}.$

Note also that

(12.13) $\quad B_{\lambda_1} = A_{\lambda_1} = R_{\lambda_1}, \qquad B_{\lambda_i} - A_{\lambda_i} = B_{\lambda_{i-1}} \qquad (i > 1).$

We now define functions $g_{\lambda_1}, \cdots, g_{\lambda_h}$, using induction, as follows:

(12.14) $$g_{\lambda_1}(x) = f_{\lambda_1}(x), \qquad x \in B_{\lambda_1},$$
and for $i > 1$,
(12.15) $$g_{\lambda_i}(x) = f_{\lambda_i}(x), \qquad x \in A_{\lambda_i},$$

(12.16) $$g_{\lambda_i}(x) = f_{\lambda_i}(x) + \phi^*{}_{\lambda_i}(x)[g_{\lambda_{i-1}}(x) - f_{\lambda_i}(x)], \qquad x \in B_{\lambda_i} - A_{\lambda_i}.$$

Because of (12.11), $g_{\lambda_i} = f_{\lambda_i}$ in R'_{λ_i}. It follows that g_{λ_i} is of class C^r in R'_{λ_i}. The same holds in $B_{\lambda_i} - A_{\lambda_i}$, because of (12.16); hence it is of class C^r in B_{λ_i}.

We now prove an inequality. Take any $x \in G$. Let

(12.17) $$x \in R_{\mu_1} \cap \cdots \cap R_{\mu_k}, \qquad \mu_1 > \cdots > \mu_k,$$

and suppose that x is in no other R_j (R_j is defined only for $j = \lambda_1, \cdots, \lambda_h$). Then for $i = 1, \cdots, k$,

(12.18) $$|D_\alpha[g_{\mu_i}(x) - f_{\mu_i}(x)]| \leqq \tfrac{1}{2} a^{\mu_i} \rho_{x\mu_i}{}^{r-[\alpha]} \epsilon(x).$$

Take first $i = 1$. Since

$$B_{\mu_1} - A_{\mu_1} = \sum_{\lambda > \mu_1} A_\lambda \subset \sum_{\lambda > \mu_1} R_\lambda,$$

x is not in $B_{\mu_1} - A_{\mu_1}$. Also $x \in R_{\mu_1} \subset B_{\mu_1}$. Hence $x \in A_{\mu_1} \subset R'_{\mu_1}$, and the left hand side of (12.18) is 0.

Now take any $i > 1$. Since the relation is trivial for $x \in R'_{\mu_i}$, we may suppose $x \in R_{\mu_i} - R'_{\mu_i}$. Using induction, and the facts $\mu_{i-1} > \mu_i$, $\rho_{x\mu_{i-1}} \leqq \rho_{x\mu_i}$ (see (12.12)), we have

$$\begin{aligned} |D_\alpha[g_{\mu_{i-1}}(x) - f_{\mu_{i-1}}(x)]| &\leqq \tfrac{1}{2} a^{\mu_{i-1}} \rho_{x\mu_{i-1}}{}^{r-[\alpha]} \epsilon(x) \\ &\leqq \tfrac{1}{2} a^{\mu_i+1} \rho_{x\mu_i}{}^{r-[\alpha]} \epsilon(x) \end{aligned}$$

Next, (12.7) and (12.12) give

$$|D_\alpha[f_{\mu_{i-1}}(x) - F(x)]| \leqq \tfrac{1}{2} a^{\mu_{i-1}+1} \bar{\rho}_{x\mu_{i-1}}{}^{r-[\alpha]} \epsilon(x) \leqq \tfrac{1}{2} a^{\mu_i+1} \rho_{x\mu_i}{}^{r-[\alpha]} \epsilon(x).$$

Also, using (12.8),

$$|D_\alpha[f_{\mu_i}(x) - F(x)]| \leqq \tfrac{1}{2} a^{\mu_i+1} \rho_{x\mu_i}{}^{r-[\alpha]} \epsilon(x).$$

These three relations give

(12.19) $$|D_\alpha[g_{\mu_{i-1}}(x) - f_{\mu_i}(x)]| \leqq \tfrac{3}{2} a^{\mu_i+1} \rho_{x\mu_i}{}^{r-[\alpha]} \epsilon(x).$$

From (12.16) and (12.10) we have

(12.20) $$g_{\lambda_j}(y) = g_{\lambda_{j-1}}(y), \qquad y \in B_{\lambda_j} - R_{\lambda_j}.$$

Since x is in no R_{λ_j} with $\mu_{i-1} > \lambda_j > \mu_i$, it follows, if $\mu_i = \lambda_m$, that $g_{\lambda_{m-1}}(x) = g_{\mu_{i-1}}(x)$. Putting this in (12.16), with i replaced by m, and differentiating gives

$$D_\alpha[g_{\mu_i}(x) - f_{\mu_i}(x)] = \sum_\beta \binom{\alpha}{\beta} D_\beta \phi^*{}_{\mu_i}(x) D_{\alpha-\beta}[g_{\mu_{i-1}}(x) - f_{\mu_i}(x)].$$

As in **11**, there are at most c terms in the sum in (12.9), for fixed x. Applying (10.2) and (12.19) and recalling that $\rho(x, P) = \rho_{x\mu_i}$ here, we find (as in **11**)

$$| D_\alpha[g_{\mu_i}(x) - f_{\mu_i}(x)] | \leqq c[(r+1)!]^n (21 n^{1/2})^r N \tfrac{3}{2} \rho_{x\mu_i}{}^{r-[\alpha]} a^{\mu_i+1} \epsilon(x);$$

using (12.2) gives (12.18).

Noting that $B_{\lambda_h} = G$, our required function is

$$f^*(x) = g_{\lambda_h}(x), \qquad x \,\varepsilon\, G. \tag{12.21}$$

We must show that

$$| D_\alpha[f^*(x) - F(x)] | \leqq \epsilon(x), \qquad x \,\varepsilon\, G, \; [\alpha] \leqq r. \tag{12.22}$$

Take any $x \,\varepsilon\, G$; apply (12.17) and (12.18). Since x is in no R_{λ_i} with $\lambda_i < \mu_k$, (12.20) gives $g_{\mu_k}(x) = g_{\lambda_h}(x) = f^*(x)$. Choose $y \,\varepsilon\, A_{\mu_k}$ so that $\rho(x, y) < \zeta_{\mu_k}(y)$. Then

$$\rho_{x\mu_k} \leqq \rho(x, y) < \zeta_{\mu_k}(y) = \eta_{\mu_k}(y) < 1.$$

Also $a < 1$. Hence, using (12.18) with $i = k$, we have

$$| D_\alpha[f^*(x) - f_{\mu_k}(x)] | \leqq \tfrac{1}{2}\epsilon(x).$$

Combining this with (12.6) (with $\lambda_i = \mu_k$) gives (12.22).

13. Completion of the proof of Theorem 2. We must still show that f^* is in I^*. To this end, we shall replace the functions f_{λ_i} and g_{λ_i} by functions f'_{λ_i} and g'_{λ_i} in I^*. To begin with, for each i, let ψ_{λ_i} be a function of class C^r in G such that

$$\psi_{\lambda_i}(x) = 1, \qquad x \,\varepsilon\, G - \sum_{j>i} R'_{\lambda_j}, \tag{13.1}$$

$$\psi_{\lambda_i}(x) = 0, \qquad x \,\varepsilon\, R''_{\lambda_i}, \tag{13.2}$$

where R''_{λ_i} is a neighborhood of $\sum_{j>i} A_{\lambda_j} = G - B_{\lambda_i}$ in G. (We may construct such a function for instance by using the standard subdivision of B_{λ_i}, and adding together all ϕ_ν such that the corresponding J_ν does not lie entirely

in some R'_{λ_j} with $j > i$.) We may clearly take $R''_{\lambda_k} \subset R''_{\lambda_j}$ for $k > j$. We now set

$$f'_{\lambda_i}(x) = \psi_{\lambda_i}(x) f_{\lambda_i}(x), \qquad x \varepsilon B_{\lambda_i}, \tag{13.3}$$

$$f'_{\lambda_i}(x) = 0, \qquad x \varepsilon G - B_{\lambda_i}. \tag{13.4}$$

Since $f'_{\lambda_i} = 0$ in R''_{λ_i}, f'_{λ_i} is of class C^r in G. Because of (13.1),

$$f'_{\lambda_i}(x) = f_{\lambda_i}(x), \qquad x \varepsilon G - \sum_{j>i} R'_{\lambda_j}. \tag{13.5}$$

We shall show that each f'_{λ_i} is in I^*. We cannot use the expression (11.10) for f_{λ_i}, since it would not satisfy condition (b) of **2** at the points of A'_{λ_i}. This, however, is easily remedied. Because of (a), **11**, each $z \varepsilon A'_{\lambda_i}$ is in a neighborhood U_z such that for each $y^\nu \varepsilon U_z$, $x^\nu = z$ (see **10**). Let $\Sigma^{(z)}$ denote the sum in (11.10) over those terms $\phi_\nu f_{x^\nu}$ such that the corresponding J_ν lies entirely in U_z. Set

$$\bar{f}_z(x) = \Sigma^{(z)} \phi_\nu(x) f_{x^\nu}(x) = [\Sigma^{(z)} \phi_\nu(x)] f_z(x), \qquad x \varepsilon B_{\lambda_i} - z, \tag{13.6}$$

and $\bar{f}_z(z) = f_z(z) = F(z)$. Since $\Sigma^{(z)} \phi_\nu \equiv 1$ in a neighborhood of z (excepting z) and $\equiv 0$ outside U_z, we may set $\bar{f}_z(x) = 0$ in $G - B_{\lambda_i}$, giving a function $\bar{f}_z$ in I (since f_z is in I).

If we set $\phi'_\nu(x) = \phi_\nu(x) f_{x^\nu}(x)$ in B_{λ_i} and $= 0$ in $G - B_{\lambda_i}$, the functions ϕ'_ν are defined and of class C^r in G, and are in I.

Let $z^1, z^2, \cdots$ denote the points of A'_{λ_i}. We may suppose the U_{z^m} have no common points. If Σ' denotes the sum of those terms corresponding to no z^m, then

$$f_{\lambda_i}(x) = \Sigma' \phi'_\nu(x) + \sum_m \bar{f}_{z^m}(x) \qquad x \varepsilon B_{\lambda_i}. \tag{13.7}$$

Arrange the functions $\psi_{\lambda_i}\phi'_\nu$, $\psi_{\lambda_i}\bar{f}_{z^m}$, using the ν appearing in Σ', in a sequence; call these functions $f^*_{\lambda_i 1}, f^*_{\lambda_i 2}, \cdots$. Then

$$f'_{\lambda_i}(x) = \sum_k f^*_{\lambda_i k}(x), \qquad x \varepsilon G; \tag{13.8}$$

this follows from (13.7) and (13.3) for $x \varepsilon B_{\lambda_i}$, and from (13.2) and (13.4) for $x \varepsilon G - B_{\lambda_i}$.

The functions $f^*_{\lambda_i k}$ are in I, and (a), **2**, holds; we shall prove (b). Take any $x \varepsilon G$. If first $x \varepsilon B_{\lambda_i}$, it is clear from the definition of the functions ϕ'_ν and $\bar{f}_{z^m}$ that at most a finite number of the functions ϕ'_ν (ν in Σ'), $\bar{f}_{z^m}$, are $\not\equiv 0$ in some neighborhood of x; hence this holds also for the $f^*_{\lambda_i k}$. If next $x \varepsilon G - B_{\lambda_i}$, then $x \varepsilon R''_{\lambda_i}$, and (13.2) shows that all the $f^*_{\lambda_i k}$ vanish in R''_{λ_i}. Thus (b) holds, and we have shown that f'_{λ_i} is in I^*.

We now define the g'_{λ_i}. Set $\phi^*_{\lambda_i}(x) = 0$ in A_{λ_i}. From (12.11) we see that $\phi^*_{\lambda_i}$ is now of class C^r in B_{λ_i}. Define

$$(13.9) \qquad g'_{\lambda_1}(x) = f'_{\lambda_1}(x), \qquad x \in G,$$

and for $i > 1$,

$$(13.10) \qquad g'_{\lambda_i}(x) = f'_{\lambda_i}(x) + \phi^*_{\lambda_i}(x)[g'_{\lambda_{i-1}}(x) - f'_{\lambda_i}(x)], \qquad x \in B_{\lambda_i},$$

and $g'_{\lambda_i}(x) = 0$, $x \in G - B_{\lambda_i}$. We see at once that $g'_{\lambda_i}(x) = 0$ in R''_{λ_i}; it follows that g'_{λ_i} is of class C^r in G.

Let us show that

$$(13.11) \qquad g'_{\lambda_i}(x) = g_{\lambda_i}(x), \qquad x \in G - \sum_{j>i} R'_{\lambda_j}.$$

For $i = 1$, this follows from (13.9), (13.5) and (12.14). Now take $i > 1$. If $x \in A_{\lambda_i}$, $\phi^*_{\lambda_i}(x) = 0$, and (13.10), (13.5) and (12.15) apply. Now take $x \in B_{\lambda_i} - A_{\lambda_i}$. If first $x \in R'_{\lambda_i}$, then (13.10), (12.11), (13.5) and (12.16) give the result. If not, then $x \in B_{\lambda_i} - \sum_{j>i-1} R'_{\lambda_j}$. Using induction gives $g'_{\lambda_{i-1}}(x) = g_{\lambda_{i-1}}(x)$, and (13.11) now follows from (13.10), (13.5) and (12.16).

We now prove that each g'_{λ_i} is in I^*. For $i = 1$, it follows from (13.9). For $i > 1$, we use induction. Note that if we replace $\phi^*_{\lambda_i}$ by a function $\phi^{**}_{\lambda_i}$ which equals $\phi^*_{\lambda_i}$ in $G - R''_{\lambda_i}$ and which goes to zero together with derivatives as we approach $G - B_{\lambda_i}$, we may set $\phi^{**}_{\lambda_i} = 0$ in $G - B_{\lambda_i}$, making it of class C^r in G, and we may use $\phi^{**}_{\lambda_i}$ in place of $\phi^*_{\lambda_i}$ in (13.10), since $g'_{\lambda_{i-1}} = f'_{\lambda_i} = 0$ in R''_{λ_i}. Now (13.10), with $\phi^{**}_{\lambda_i}$, holds in G. Using this, and the facts that f'_{λ_i} and $g'_{\lambda_{i-1}}$ are in I^* and that I^* is an ideal, shows that g'_{λ_i} is in I^*.

Finally, setting $i = h$ in (13.11) and using (12.21) gives

$$(13.12) \qquad f^*(x) = g_{\lambda_h}(x) = g'_{\lambda_h}(x), \qquad x \in G.$$

Since g'_{λ_h} is in I^*, f^* is in I^*, and the proof of Theorem 2 is complete.

14. Restrictions on the local ideals. We consider here the problem: What sets of local r-ideals are the local ideals of an r-ideal? A good answer to the question seems quite difficult to give, especially for $r \geqq 2$. For a better understanding of the problem, we shall give two theorems, which do hardly more than show wherein the difficulty lies.

The first theorem says that each r-value of each local ideal must be realized by a function, each of whose r-values lies in the corresponding local ideal.

THEOREM 4. *Given a set of local r-ideals $H(x)$ for each x in an open subset G of E^n, there is an r-ideal I in G such that*

$$\mathfrak{J}_r(I, x) = H(x), \qquad x \in G, \tag{14.1}$$

if and only if the following holds. For each x^0 in G and for each r-value $a^0 \in H(x^0)$ there exists a function f of class C^r in G such that

$$V_r(f, x^0) = a^0; \qquad V_r(f, x) \in H(x), \qquad x \in G. \tag{14.2}$$

To prove the necessity, suppose that I exists. Given $a^0 \in H(x^0)$, since $\mathfrak{J}_r(I, x^0) = H(x^0)$, there is a function f in I such that $V_r(f, x^0) = a^0$; clearly (14.2) holds.

To prove the sufficiency, suppose the condition satisfied. Let I consist of all functions f of class C^r in G such that $V_r(f, x) \in H(x)$ for all x. Since the $H(x)$ are local ideals, I is clearly an ideal. Certainly $\mathfrak{J}_r(I, x) \subset H(x)$, all x. To prove $H(x) \subset \mathfrak{J}_r(I, x)$, all x, take any x^0 and any $a^0 \in H(x^0)$. Choose f so as to satisfy (14.2). Then f is in I, and $V_r(f, x^0) = a^0$, so that $H(x^0) \subset \mathfrak{J}_r(I, x^0)$, completing the proof.

The next theorem is obtained by applying the last theorem, not to all points of G, but to all the points where the local ideals $H(x)$ are non-trivial. Recall (**6, 7**) that the unit local r-ideal $\mathfrak{J}_r^1(x)$ consists of all r-values at x, and that for any other local ideal $H(x)$, $a_{0\ldots0} = 0$ for each r-value $a \in H(x)$, that is, for any f with $V_r(f, x) \in H(x)$, $f(x) = 0$. Thus the set A of points x where $H(x) \neq \mathfrak{J}_r^1(x)$ is the set of points at which any function f in any corresponding ideal is required to vanish. The following theorem gives conditions on the local ideals at points of A; there are of course no conditions at other points of G.

THEOREM 5. *Given the $H(x)$, there is a corresponding I as in Theorem 3 if and only if the following holds. Let A be the set of points x in G for which $H(x) \neq \mathfrak{J}_r^1(x)$; we assume A is closed. Then for each x^0 in A and for each $a^0 \in H(x^0)$ there is a neighborhood $U \subset G$ of x^0 and there are r-values $a(x) \in H(x)$ for $x \in U \cap A$ such that $a(x^0) = a^0$, and so that the following is true. For each pair of points x, x' in $U \cap A$, and for each $\alpha = (\alpha_1, \cdots, \alpha_n)$, $[\alpha] \leqq r$, define $R_\alpha(x', x)$ by the relation*

$$a_\alpha(x') = \sum_{[\beta] \leqq r-[\alpha]} (a_{\alpha+\beta}(x)/\beta!)(x'-x)^\beta + R_\alpha(x', x). \tag{14.3}$$

Then for each $x^* \in U \cap A$ and $\epsilon > 0$ there is a $\delta > 0$ such that

$$|R_\alpha(x', x)| \leqq \rho_{xx'}{}^{r-[\alpha]}\epsilon \text{ if } \rho_{xx^*}, \rho_{x'x^*} < \delta, \qquad x, x' \in U \cap A, [\alpha] \leqq r. \tag{14.4}$$

EXAMPLE. Suppose that A consists of isolated points only. Then the condition is obviously always true.

First suppose that I exists. Given x^0 and $a^0 \varepsilon H(x^0)$, choose f in I so that $V_r(f, x^0) = a^0$, and set $a(x) = V_r(f, x)$, $x \varepsilon G$. Then the R_α are simply the remainders in Taylor's formula, and (14.4) is well known.

Now suppose that the condition holds. By Theorem I of AE, we may extend the values of the $a_\alpha(x)$ $([\alpha] \leqq r)$ through U so that $a_0(x) = a_{0\ldots 0}(x)$ is of class C^r in U in terms of the $a_\alpha(x)$. (Though Theorem I of AE is stated for E only, it clearly holds for open sets U. Actually, we could apply Theorem III of AE directly.) Since $H(x) = \mathfrak{J}_r^{\,1}(x)$, $x \varepsilon U - A$, $a(x) \varepsilon H(x)$ for such x also. Let $\phi(x)$ be a function of class C^r in G which is $\equiv 0$ outside a neighborhood U' of x^0 with $\bar{U}' \subset U$, and is $\equiv 1$ in a neighborhood U'' of x^0; set

$$f(x) = \phi(x) a_0(x) \text{ in } U, \qquad f(x) = 0 \text{ in } G - U.$$

Then f is of class C^r in G, and since $a(x) \varepsilon H(x)$ $(x \varepsilon U)$ and the $H(x)$ are local ideals, $V_r(f, x) \varepsilon H(x)$ for all $x \varepsilon G$. Clearly $V_r(f, x^0) = a(x^0) = a^0$; thus the condition of the last theorem is proved, and the present theorem follows.

Remark. In the condition of the last theorem, we could drop the hypothesis that A is closed, if we assume that (14.4) holds for all x^* in U instead of all x^* in $U \cap A$; for we could then prove that A is closed (or we could use Theorem 1 of H. Whitney, "Differentiable functions defined in arbitrary subsets of Euclidean space," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 307-317).

HARVARD UNIVERSITY.

ON TOPOLOGICAL GROUPS OF HOMEOMORPHISMS.*

By Jean Dieudonné.

1. **Introduction.** The aim of this paper is to systematize and extend a number of results on topologies for homeomorphism groups which have been recently published by several authors (see [4], [5], [6], [8], [9] in the list at the end of this paper); we have tried in each case to get rid of superfluous hypotheses in the formulation of the theorems we give, and to reach the closest approximation possible to their exact range of validity. The questions treated fall into two quite distinct chapters: in the first one, we consider either the *whole* group of homeomorphisms of a uniform space, or subgroups formed of those homeomorphisms which satisfy certain mild conditions of *uniform continuity*; the main problem is to decide when a topology (defined by natural conditions of "uniform convergence" on certain subsets) is *compatible* with the group, that is, makes the product uv and the inverse u^{-1} into continuous functions; we obtain conditions (see Theorems 1, 2 and 3) which generalize previous results by G. Birkhoff [5] and R. Arens [4], the extension consisting principally in the possibility of dropping in certain cases restrictions of compactness (or local compactness) for the underlying space, assumed by these authors; in particular, we are thus able to obtain a topology which is compatible with the group of all linear automorphisms of a Banach space (Theorem 3).

In the second chapter, we restrict considerably, on the contrary, the subgroups of homeomorphisms we study, by submitting them to conditions of *equicontinuity*. The chief question here is to find conditions under which a subgroup of homeomorphisms is *locally compact*, or contained in such a subgroup; conditions for a group of homeomorphisms to be compact had already been given by S. B. Myers [9] and the author [8], and R. Arens had studied a special type of locally compact groups, those which are equicontinuous [4, p. 604]; we solve here the general problem (Theorem 5) which offers new difficulties as regards the way in which the group must be "completed" so as to become locally compact; in that result also we are able to do away with restrictions of compactness or local compactness of the underlying space.

* Received July 12, 1947.

CHAPTER I.

2. Preliminaries and notations. We follow, in this paper, the topological terminology and general notations of N. Bourbaki's treatise ([1] and [2]). All spaces considered are *uniform spaces* [Bourbaki 1, p. 85 and 89] in which every point is a closed subspace (which is equivalent to saying that the topology of the uniform space is a Hausdorff topology). Let E be such a space, $\mathfrak{A}$ the filter base of symmetrical neighborhoods of the diagonal Δ in $E \times E$ which defines the uniform structure of E; we shall denote by $\mathcal{F}(E)$ the set of all functions defined in E with values in E, by $\mathcal{C}(E)$ (resp. $\mathcal{C}^*(E)$) the subset of $\mathcal{F}(E)$ consisting of all continuous (resp. uniformly continuous) functions. Let Σ be a family of closed sets in E, such that every union of a finite number of sets from Σ belongs to Σ; for every $A \in \Sigma$ and every $V \in \mathfrak{A}$, let $W(A, V)$ be the set of all couples (u, v) of elements of $\mathcal{F}(E)$ such that $(u(x), v(x)) \in V$ for every $x \in A$. It is easily seen that the $W(A, V)$ form the base of a filter of subsets of $\mathcal{F}(E) \times \mathcal{F}(E)$ which defines on $\mathcal{F}(E)$ a uniform structure, which we shall denote by $\mathcal{U}_\Sigma$, the corresponding topology on $\mathcal{F}(E)$ being written $\mathcal{T}_\Sigma$. We shall always suppose that every point of E belongs to at least one set of Σ, in which case $\mathcal{T}_\Sigma$ is a Hausdorff topology. The three most important cases to be considered are the following:

1° Σ consists of the single set E;

2° Σ consists of the compact subsets of E;

3° Σ consists of the finite subsets of E.

The uniform structures in these three cases will be written $\mathcal{U}_u$, $\mathcal{U}_c$, $\mathcal{U}_s$, respectively, the corresponding topologies $\mathcal{T}_u$, $\mathcal{T}_c$, $\mathcal{T}_s$; convergence in $\mathcal{T}_u$ means uniform convergence on E; convergence in $\mathcal{T}_c$, uniform convergence on every compact subset of E (see [3] and [9]); convergence in $\mathcal{T}_s$ is pointwise convergence. When $\mathcal{F}(E)$ is given the uniform structure $\mathcal{U}_u$, $\mathcal{U}_c$, or $\mathcal{U}_s$, the uniform space thus defined will be written $\mathcal{F}_u(E)$, $\mathcal{F}_c(E)$, $\mathcal{F}_s(E)$ respectively; $\mathcal{C}_u(E)$, $\mathcal{C}_c(E)$, $\mathcal{C}_s(E)$ will denote its subspace $\mathcal{C}(E)$ (with the same uniform structure). The uniform space $\mathcal{F}_s(E)$ is identical with the product space E^E [Bourbaki 1, p. 118]; each of the structures $\mathcal{U}_u$, $\mathcal{U}_c$, $\mathcal{U}_s$ is finer [Bourbaki 1, p. 88] than the next; if E is compact, $\mathcal{U}_u = \mathcal{U}_c$; if E is discrete, $\mathcal{U}_c = \mathcal{U}_s$.

We state without proof the following properties, which are either known or easy generalizations of classical results:

LEMMA 1. *If E is complete, so is $\mathcal{F}(E)$ for the uniform structure $\mathcal{U}_\Sigma$.*

LEMMA 2. *$\mathfrak{H}(E)$ and $\mathfrak{H}^*(E)$ are closed subspaces in $\mathfrak{F}_u(E)$, and therefore complete if E is complete.*

LEMMA 3. *If E is locally compact, $\mathfrak{H}(E)$ is a closed subspace in $\mathfrak{F}_c(E)$.*

In relation with the space $\mathfrak{H}_c(E)$, we recall that its topology $\mathfrak{T}_c$ is *independent* of the uniform structure of E, and can be defined in terms of the topology of E alone (see [3], p. 489).

LEMMA 4. *If Σ is such that every point of E is interior to at least one set of Σ, then the application $(x, u) \rightarrow u(x)$ of $E \times \mathfrak{H}(E)$ into E is continuous when $\mathfrak{H}(E)$ is given the topology $\mathfrak{T}_\Sigma$.*

3. Continuity of $(u, v) \rightarrow uv$. For any two functions u, v belonging to $\mathfrak{F}(E)$, uv will always be the function $x \rightarrow u(v(x))$; if u is a one-one application of E on itself, u^{-1} will be the reciprocal function. We are interested first in getting a sufficient condition for the application $(u, v) \rightarrow uv$ of $\mathfrak{H}(E) \times \mathfrak{H}(E)$ into $\mathfrak{H}(E)$ to be continuous at a given point (u_0, v_0), when $\mathfrak{H}(E)$ is given the topology $\mathfrak{T}_\Sigma$.

PROPOSITION 1. *If, for arbitrary $A \varepsilon \Sigma$, there exist $B \varepsilon \Sigma$ and $V \varepsilon \mathfrak{U}$ such that:*

1° $V(v_0(A)) \subset B$;

2° *u_0 is uniformly continuous in B;*

then uv is a continuous function of (u, v) at the point (u_0, v_0) for the topology $\mathfrak{T}_\Sigma$.

We have to show that, given arbitrarily $A \varepsilon \Sigma$ and $W \varepsilon \mathfrak{U}$, it is possible to find $B \varepsilon \Sigma$, $C \varepsilon \Sigma$, $U \varepsilon \mathfrak{U}$ and $V \varepsilon \mathfrak{U}$ such that the conditions $(u(y), u_0(y)) \varepsilon U$ for every $y \varepsilon B$ and $(v(z), v_0(z)) \varepsilon V$ for every $z \varepsilon C$ insure that $(u(v(x)), u_0(v_0(x))) \varepsilon W$ for every $x \varepsilon A$. Take $C = A$ and U such that $U^2 \subset W$; then it is enough to show that $B \varepsilon \Sigma$ and $V \varepsilon \mathfrak{U}$ may be found such that, in the first place, $(v(x), v_0(x)) \varepsilon V$ for $x \varepsilon A$ implies $(u_0(v(x)), u_0(v_0(x))) \varepsilon U$ for $x \varepsilon A$, and on the other hand, $(u(y), u_0(y)) \varepsilon U$ for $y \varepsilon B$ implies $(u(v(x)), u_0(v(x))) \varepsilon U$ for $x \varepsilon A$. Let $V_0 \varepsilon \mathfrak{U}$ and $B \varepsilon \Sigma$ be taken such that $V_0(v_0(A)) \subset B$ and that u_0 is uniformly continuous on B; then $(v(x), v_0(x)) \varepsilon V_0$ in A will imply $v(A) \subset V_0(v_0(A)) \subset B$; therefore $(u(y), u_0(y)) \varepsilon U$ for $y \varepsilon B$ implies $(u(v(x)), u_0(v(x))) \varepsilon U$ for $x \varepsilon A$; on the other hand, as u_0 is uniformly continuous in B, a $V \varepsilon \mathfrak{U}$ may be found such that $V \subset V_0$ and $(v(x), v_0(x)) \varepsilon V$ for $x \varepsilon A$ implies $(u_0(x)), u_0(v_0(x))) \varepsilon U$ for $x \varepsilon A$, which completes the proof.

PROPOSITION 2. *If u_0 is uniformly continuous in E, $(u, v) \to uv$ is an application of $\mathscr{B}_u(E) \times \mathscr{B}_u(E)$ into $\mathscr{B}_u(E)$ which is continuous at the point (u_0, v_0).*

PROPOSITION 3. *If E is locally compact, $(u, v) \to uv$ is a continuous application of $\mathscr{B}_c(E) \times \mathscr{B}_c(E)$ into $\mathscr{B}_c(E)$.*

Proposition 2 is an obvious corollary of Proposition 1. To prove Proposition 3, let A be any compact subset of E; $v_0(A)$ is also compact, and, since E is locally compact, $v_0(A)$ has a compact neighborhood B; the conditions of Proposition 1 are therefore satisfied [Bourbaki 1, p. 109, Th. 2 and p. 111, Prop. 1].

4. Continuity of $u \to u^{-1}$. In this section, we consider only the subset H of $\mathscr{B}(E)$ consisting of the *homeomorphisms* of E on itself; u^{-1} is therefore defined in H. We want to find conditions for continuity of the application $u \to u^{-1}$ of H on itself at a given point u_0, H being given the topology $\mathcal{T}_\Sigma$.

PROPOSITION 4. *If there is a subfamily Σ' of Σ such that every set of Σ is contained in a finite union of sets of Σ', and if given arbitrarily $A \in \Sigma'$, there exists $B \in \Sigma'$ and $V \in \mathfrak{A}$ such that:*

1° *u_0^{-1} is uniformly continuous in $V(A)$;*

2° *the relations $(u_0(x), u(x)) \in V$ (u in H) for every $x \in B$ implies $A \subset u(B)$;*

then u^{-1} is a continuous function of u at the point u_0 for the topology $\mathcal{T}_\Sigma$.

We have to prove that, given arbitrarily $A \in \Sigma$ and $W \in \mathfrak{A}$, it is possible to find $B \in \Sigma$ and $V \in \mathfrak{A}$ such that the condition $(u_0(x), u(x)) \in V$ for $x \in B$ implies $(u_0^{-1}(y), u^{-1}(y)) \in W$ for $y \in A$. If A is contained in the union of a finite number of sets $A_i \in \Sigma'$, and, for each i, $B_i \in \Sigma$ and $V_i \in \mathfrak{A}$ are such that $(u_0(x), u(x)) \in V_i$ for $x \in B_i$ implies $(u_0^{-1}(y), u^{-1}(y)) \in W$ for $y \in A_i$, then $B = \bigcup_i B_i$ and $V = \bigcap_i V_i$ will fulfill the required conditions; we may therefore limit ourselves to the case when $A \in \Sigma'$. Let $V_0 \in \mathfrak{A}$ and $B \in \Sigma$ be such that $(u_0(x), u(x)) \in V_0$ for $x \in B$ implies $A \subset u(B)$ (u in H), and that u_0^{-1} is uniformly continuous in $V_0(A)$; there exists therefore a $V \in \mathfrak{A}$ such that $V \subset V_0$ and that the relation $(u_0(u^{-1}(y)); y) \in V$ for $y \in A$ implies $(u^{-1}(y), u_0^{-1}(y)) \in W$; but if we put $x = u^{-1}(y)$, the relation $(u_0(u^{-1}(y)), y) \in V$ can be written $(u_0(x), u(x)) \in V$; if it holds in B, it will hold *a fortiori* in $u^{-1}(A) \subset B$, and the proof is thus completed.

PROPOSITION 5. *If u_0 is a homeomorphism of E upon itself such that u_0^{-1} is uniformly continuous in E, $u \rightarrow u^{-1}$ is an application of H upon itself which is continuous at the point u_0 for the topology $\mathcal{T}_u$.*

In fact, for the topology $\mathcal{T}_u$, Σ consists of a single set E, and the second condition of Proposition 4 is obviously satisfied since $u(E) = E$ for every homeomorphism $u \varepsilon H$.

Except in that particular case of the topology $\mathcal{T}_u$, it seems necessary to impose rather heavy restrictions on the space E or the family of subsets Σ in order to satisfy condition 2° of Proposition 4. It is here that the notion of *connection* plays an important part, as was first pointed out by G. Birkhoff ([5], p. 871). We have in fact the following general property:

PROPOSITION 6. *Suppose that there exists a subfamily Σ' of Σ such that every set of Σ is contained in a finite union of sets of Σ', and that, for every set $A \varepsilon \Sigma'$, there exists $B \varepsilon \Sigma$ and $V \varepsilon \mathfrak{A}$ such that:*

1° *u_0^{-1} is uniformly continuous in $V(A)$;*

2° *there exists a connected set C such that $V(A)$ is contained in C and $V(C)$ contained in the interior of $u_0(B)$.*

Then u^{-1} is a continuous function of u at the point $u_0 \varepsilon H$ for the topology $\mathcal{T}_\Sigma$.

We have only to show, according to Proposition 4, that for every $u \varepsilon H$ such that $(u_0(x), u(x)) \varepsilon V$ for every $x \varepsilon B$, we have $A \subset u(B)$. Notice first that $u(u_0^{-1}(A)) \subset V(A) \subset C$; therefore $u^{-1}(C)$ contains points of B; suppose that $u^{-1}(C)$ is not contained in B; as C is connected, so is $u^{-1}(C)$; that set, having points in B and points not belonging to B, would contain at least one point z belonging to the frontier of B. But then $u_0(z)$ belongs to the frontier of $u_0(B)$, and $u(z)$ belongs to C; the assumptions make it impossible that one should have $(u_0(z), u(z)) \varepsilon V$, which is a contradiction.

5. Topologies compatible with groups of homeomorphisms. Propositions 2 and 5 yield immediately the following result:

THEOREM 1. *Let H^* be the group of all automorphisms of the uniform structure of E* (that is, homeomorphisms u such that u and u^{-1} are both uniformly continuous). *Then the topology $\mathcal{T}_u$ is compatible with the group H^** (that is, uv and u^{-1} are continuous in $H^* \times H^*$ and H^*, respectively, for that topology).

In the particular case when E is *compact*, H^* coincides, of course, with

the *whole* group of homeomorphisms H of the space E, and Theorem 1 becomes a well known result (see A. Weil [12], p. 131 and R. Arens [4], p. 597).

In the same way, combining Propositions 3 and 6 gives the following theorem due to R. Arens [4, p. 598]:

THEOREM 2. *If E is a locally compact and locally connected space the topology $\mathfrak{T}_c$ is compatible with the group H of all homeomorphisms of E.*

We have only to take, in Proposition 6, for Σ' the family of compact connected subsets of E; for C a compact connected neighborhood of A; to take V such that $V(A) \subset C$, and that $V(C)$ is contained in the interior of a compact neighborhood K of C; and finally to take $B = u_0^{-1}(K)$, which is compact. It follows from the assumptions on E that all these choices are possible, and the conditions of Proposition 6 are then satisfied, which proves that u^{-1} is continuous in H.

Theorem 2 is an extension and precision of a previous result of G. Birkhoff [5, p. 871]; an interesting extension of that result in another direction is the following:

THEOREM 3. *Let E be a metric space, Σ the family of all "balls" $d(x_0, x) \leq r$ of center x_0 (d distance on E). Suppose that there exists an increasing sequence (r_n) of positive numbers, tending to $+\infty$, such that each ball $d(x_0, x) \leq r_n$ is connected. Then, if L is the group of all homeomorphisms of E which are bounded and uniformly continuous as well as their reciprocals on each ball of center x_0, the topology $\mathfrak{T}_\Sigma$ is compatible with the group L.*

In fact, the conditions of Propositions 1 and 6 are obviously satisfied, taking as Σ' the subfamily of all the balls $d(x_0, x) \leq r_n$.

Interesting subgroups of the group L are: the group of all *Lipschitzian* homeomorphisms of E (that is, such that $d(u(x), u(y)) \leq k \cdot d(x, y)$, k depending on u), and, in the case where E is a normed space, the still more particular group of all *linear* homeomorphisms of E.

6. Uniform structures on groups of homeomorphisms. In this section, we investigate the properties of the standard uniform structures of a topological group of homeomorphisms, (that is, the left, right and two-sided structures deduced from the topology of the group; see A. Weil [11, p. 30], and Bourbaki [2, p. 23 and 31-32]) in the cases when we have succeeded up to now in defining such a group.

In the first place, it is easily seen that on the group H^* of all auto-

morphisms of the uniform structure of E, topologized by $\mathfrak{T}_u$, the *right* uniform structure is identical with the uniform structure $\mathcal{U}_u$ of uniform convergence in E. In fact, to define that structure, for each $V \in \mathfrak{A}$, we have to take the set of all couples (u, v) of elements of H^* such that $(v(u^{-1}(x)), x) \in V$ for every $x \in E$; but as u is an homeomorphism of E, the last relation is obviously equivalent to $(v(y), u(y)) \in V$ for every $y \in E$, which proves our contention.

Unfortunately, the right uniform structure of H^* behaves in a very unpleasant manner, even when one restricts the space E in the most drastic way, taking for instance for E a compact interval. I have in fact shown in a previous Note [7] that even in that very special case, not only is H^* not complete for its right uniform structure, but it cannot even be completed; of course, the same situation prevails for the *left* uniform structure of H^*, and both structures are necessarily distinct.

It turns out that, in all three cases considered in Section **5**, the *two-sided* uniform structure has, on the contrary, a quite reasonable behavior, as the following propositions show (see R. Arens [4], p. 602),

PROPOSITION 7. *If E is a complete uniform space, the group H^* (topologized by $\mathfrak{T}_u$) is complete for its two-sided uniform structure.*

PROPOSITION 8. *If E is a locally compact and locally connected space, the group H (topologized by $\mathfrak{T}_c$) is complete for its two-sided uniform structure.*

PROPOSITION 9. *Let E, Σ and L satisfy the conditions of Theorem 3. If in addition E is complete, the group L* (topologized by $\mathfrak{T}_\Sigma$) *is complete for its two-sided uniform structure.*

The idea of investigating the two-sided uniform structure for homeomorphism groups is due to R. Arens, who proved Proposition 7 in the particular case when E is *compact* (in which case, of course, $H^* = H$, and $\mathfrak{T}_u$ is identical to $\mathfrak{T}_c$; see [4], p. 602, three first lines of proof of Theorem 6). Proposition 8 has also been proved by R. Arens, who takes advantage of the fact [4, p. 601] that when E is locally compact and locally connected, E_1 is the compact space obtained by adding to E a "point at infinity," and H_1 is the group of all homeomorphisms of E_1, with the topology $\mathfrak{T}_u$, then $\mathfrak{T}_c$ on H is identical to the topology obtained by considering H as a subgroup of H_1; Proposition 8 is thus reduced to the particular case of Proposition 7 dealing with compact spaces. We shall give a proof of Proposition 9, which does not seem to be reducible to Proposition 7 in such a way; of course, our argument will be very similar to Arens's, and we leave to the reader the easy task of

adapting it to prove Proposition 7, or to give an independent proof of Proposition 8.

Let then ϕ be a Cauchy filter on the group L for the two-sided uniform structure of that group. Let A be any ball of center x_0, K_0 a connected ball such that A is contained in the interior of K_0, K a ball such that K_0 is contained in the interior of K. There exists a $V_0 \varepsilon \mathfrak{A}$ such that $V_0(A)$ is contained in K_0, and that $V_0(K_0)$ is contained in the interior of K. From the hypothesis, it follows that there exists a set $M \varepsilon \phi$ such that, for every couple (u, v) of elements of M, one has $(u^{-1}(v(x)), x) \varepsilon V_0$ and $(u(v^{-1}(x)), x) \varepsilon V_0$ for each $x \varepsilon K$ (in that second relation we make use of the fact that the topology $\mathcal{T}_\Sigma$ is compatible with the group L, and therefore the symmetrical of any neighborhood of the identity is again such a neighborhood); this is equivalent to saying that $(u^{-1}(y), v^{-1}(y)) \varepsilon V_0$ for $v^{-1}(y) \varepsilon K$, and $(u(z), v(z)) \varepsilon V_0$ for $v(z) \varepsilon K$. Let v_0 be an element of M, and put $B = v_0(K)$, $C = v_0^{-1}(K)$. For every $u \varepsilon M$, one has $(u^{-1}(y), v_0^{-1}(y)) \varepsilon V_0$ for every $y \varepsilon B$; let us show that this implies $u^{-1}(B) \supset A$. In fact, we have for $y \varepsilon v_0(A) \subset B$, $u^{-1}(y) \varepsilon V_0(v_0^{-1}(y)) \subset V_0(A) \subset K_0$; there are, therefore, points of K_0 which belong to $u^{-1}(B)$; if A were not contained in $u^{-1}(B)$, K_0 would have points outside $u^{-1}(B)$, and therefore points on the frontier of $u^{-1}(B)$, since K_0 is connected. Such a point has the form $u^{-1}(y)$ where y belongs to the frontier of B; $v_0^{-1}(y)$ is then on the frontier of K, and the choice of K_0 makes it impossible to have $(u^{-1}(y), v_0^{-1}(y)) \varepsilon V_0$.

Now, ϕ being a Cauchy filter, and B being bounded there exists, for every $V \varepsilon \mathfrak{A}$, a set $N \varepsilon \phi$, such that $N \subset M$ and that, for every couple (u, v) of elements of N, one has $(u(v^{-1}(t)), t) \varepsilon V$ for every $t \varepsilon B$, which means $(u(x), v(x)) \varepsilon V$ for every $x \varepsilon v^{-1}(B)$; but as $N \subset M$, one has $v^{-1}(B) \supset A$, therefore $(u(x), v(x)) \varepsilon V$ for every $x \varepsilon A$, that is, ϕ converges *uniformly* on every ball A; its limit u_0 is therefore a continuous application of E into E. In the same way (but using the set C instead of B) one proves that ϕ^{-1} converges to a continuous application u'_0 of E into E. From the continuity of the product in L, it follows that uu^{-1} and $u^{-1}u$ converge (according to the filter ϕ) uniformly in every ball, towards $u_0u'_0$ and u'_0u_0 respectively; this shows that $u_0u'_0$ and u'_0u_0 are both the identity, whence it follows that u_0 is a biunivocal application of E upon itself, and u'_0 its reciprocal; as they are obviously bounded and uniformly continuous on every ball, this completes the proof of Proposition 9.

CHAPTER II.

7. Preliminary lemmas. The counterexamples given by G. Birkhoff [5, p. 871] and R. Arens [4, p. 601] leave but little hope of discovering new interesting topological groups of homeomorphisms without restricting considerably more than we have been doing in Chapter I the extension of such groups. We are led to such restrictions, on the other hand, by inquiring after topological groups of homeomorphisms which do not exhibit such awkward features as the ones we have noticed in Section **6** (with respect to the uniform structures of the groups we considered); and, in the first place, we look of course after groups of homeomorphisms which belong to the best known class of topological groups, that is, *compact* or *locally compact groups* (or subgroups of such groups). This in turn (due to the classical Ascoli theorem) leads to the consideration of subgroups of homeomorphisms which are characterized by their properties in relation with the notion of *equicontinuity*.

We recall that a subset H of $\mathscr{B}(E)$ is said to be *equicontinuous* in E if, for every $V \varepsilon \mathfrak{A}$ and every $x_0 \varepsilon E$, there exists a neighborhood U of x_0 such that for every $x \varepsilon U$ and *every* $u \varepsilon H$, one has $(u(x_0), u(x)) \varepsilon V$. H is said to be *uniformly equicontinuous* in E if, given $V \varepsilon \mathfrak{A}$, there exists $W \varepsilon \mathfrak{A}$ such that the relation $(x, y) \varepsilon W$ implies $(u(x), u(y)) \varepsilon V$ for *every* $u \varepsilon H$; of course, uniform equicontinuity implies equicontinuity, but in general both notions are distinct [1]; they are equivalent, however, when E is *compact*.

These notions admit relativizations: for a subspace $A \subset E$, we say that H is *equicontinuous* (resp. *uniformly equicontinuous*) *in* A if the restrictions of the functions of H in A form a subset which is equicontinuous (resp. uniformly equicontinuous) in the uniform space A (as applications of A into E). We shall especially consider sets H of functions which are equicontinuous *in every compact subset* of E [2]; such sets are subsets of the subset $\mathcal{P}(E)$ of $\mathcal{F}(E)$ which consist of all applications of E into E which are *continuous on every compact subset* of E [2]; Lemma 3 shows that if E is complete the space $\mathcal{P}_c(E)$ (that is, $\mathcal{P}(E)$ with the uniform structure $\mathcal{U}_c$) is *complete.*

---

[1] We point out that in his paper [4], R. Arens uses the word "equicontinuous" in the sense which we attach here to "uniformly equicontinuous"; the reader should be well aware of this difference when comparing Arens's results with ours.

[2] It is easily seen that when E is either locally compact or first countable, a function from E to E which is continuous on every compact subset of E is continuous outright in E; and a set H of such functions which is equicontinuous in every compact subset of E is equicontinuous in E.

We state again without proof a number of results on equicontinuous sets we shall need in what follows (see [3], p. 490-491 and [9]).

LEMMA 5. *If H is equicontinuous* (resp. *uniformly equicontinuous*) *in E, the closure of H in the space $\mathcal{F}_s(E)$* (that is, for pointwise convergence in E) *is equicontinuous* (resp. *uniformly equicontinuous*) *in E.*

LEMMA 6. *If H is equicontinuous in every compact subset of E, the uniform structures $\mathcal{U}_s$ and $\mathcal{U}_c$ are identical on H; the closure of H in the space $\mathcal{P}_c(E)$ is identical to its closure in the space $\mathcal{F}_s(E)$, and is equicontinuous in every compact subset of E.*

LEMMA 7. *In order that a subset H of the space $\mathcal{P}_c(E)$ be relatively compact in $\mathcal{P}_c(E)$* (that is, in order to have a compact closure in that space) *it is necessary and sufficient that:*

1° *H be equicontinuous in every compact subset of E;*

2° *for every $x \in E$, the set $H(x)$* (set of all $u(x)$ for $u \in H$) *be relatively compact in E.*

When those conditions are satisfied, for every compact subset A of E, the set $H(A)$ (set of all $u(x)$ for $u \in H$ and $x \in A$) *is relatively compact in E.*

This last lemma specializes of course to the classical Ascoli theorem when E is locally compact. It is to be stressed that in Lemma 7 the space E is *not* supposed to be complete; in fact, the topology of $\mathcal{P}_c(E)$ depending only on the topology of E, but *not* on the uniform structure of that space, the part played by that structure in the formulation of Lemma 7 is only apparent: if H is such that $H(x)$ is relatively compact for every $x \in E$, when H is equicontinuous on every compact set for *one* uniform structure on E (compatible with the topology) it has the same property for *all* such uniform structures on E; this is not to be wondered at, since, by Lemma 7, for every compact subset A of E, $H(A)$ is then compact in E, and therefore all possible uniform structures on E give on A and on $H(A)$ the *same* uniform structures [Bourbaki 1, p. 107].

8. Equicontinuous groups of homeomorphisms.

THEOREM 4. *If H is a group of homeomorphisms of E which is equicontinuous in E, the topology $\mathcal{F}_s$* (identical on H with $\mathcal{F}_c$, according to Lemma 6) *is compatible with the group H.*

We have to prove that uv and u^{-1} are continuous in $H \times H$ and H respectively, for the topology $\mathcal{F}_s$.

1° *Continuity of uv.* As the space $\mathcal{F}_s(E)$ is identical with the product space E^E, to prove that $(u, v) \to uv$ is continuous, it is enough to show that, for any $x_0 \in E$, the application $(u, v) \to u(v(x_0))$ of $H \times H$ into E is continuous [Bourbaki 1, p. 44]. Let u_0 and v_0 be any two points of H, and $y_0 = v_0(x_0)$; for any $V \in \mathfrak{A}$, there exists a $U \in \mathfrak{A}$ such that the relation $(y, y_0) \in U$ implies $(u(y), u(y_0)) \in V$ for *every* $u \in H$. Let $v \in H$ be such that $(v(x_0), v_0(x_0)) \in U$, $u \in H$ be such that $(u(y_0), u_0(y_0)) \in V$; we have $(u(v(x_0)), u(v_0(x_0))) \in V$, and $(u(v_0(x_0)), u_0(v_0(x_0))) \in V$, whence $(u(v(x_0)), u_0(v_0(x_0))) \in V^2$, which proves our assertion.

2° *Continuity of* u^{-1}. For the same reason, it is enough to show that, for any $x_0 \in E$, $u \to u^{-1}(x_0)$ is a continuous application of H into E. Let u_0 be any point of H, and $y_0 = u_0^{-1}(x_0)$; for any $V \in \mathfrak{A}$, there exists $U \in \mathfrak{A}$ such that $(x, x_0) \in U$ implies $(u^{-1}(x), u^{-1}(x_0)) \in V$ for *every* $u \in H$. Let $u \in H$ be such that $(u(y_0), u_0(y_0)) \in U$, that is $(u(y_0), x_0) \in U$; we have therefore $(y_0, u^{-1}(x_0)) \in V$, that is $(u_0^{-1}(x_0), u^{-1}(x_0)) \in V$, which completes the proof.

9. Uniform structures on equicontinuous groups of homeomorphisms.

PROPOSITION 10. *Let H be a group of homeomorphisms of E, equicontinuous in E, and made into a topological group by the topology $\mathcal{F}_s$ on H. The left uniform structure of that group is then finer than the structure $\mathcal{U}_s$ on H; both structures are identical when H is uniformly equicontinuous in E.*

In fact, to form a base of the filter of neighborhoods of the diagonal in $H \times H$, which defines the left uniform structure of that topological group, we must take, for an arbitrary $U \in \mathfrak{A}$ and an arbitrary finite subset $(x_i)_{1 \leq i \leq n}$ of E, the set T of all couples (u, v) of $H \times H$ such that $(u^{-1}(v(x_i)), x_i) \in U$ for $1 \leq i \leq n$. As H is equicontinuous, for every $V \in \mathfrak{A}$, there exists a $U \in \mathfrak{A}$ such that each of the n relations $(x, x_i) \in U$ implies, for every $u \in H$, $(u(x), u(x_i)) \in V$; if U is chosen in that way, for every couple $(u, v) \in T$, one has $(v(x_i), u(x_i)) \in V$ for every i; therefore, the intersection of $H \times H$ with the set $W(x_1, \cdots, x_n, V)$ contains T, which proves the first part of proposition 10.

Let now H be uniformly equicontinuous; for every $U \in \mathfrak{A}$, there exists a $V \in \mathfrak{A}$ such that $(x, y) \in V$ implies $(u(x), u(y)) \in U$ for every $u \in H$; therefore, for each i, the relation $(v(x_i), u(x_i)) \in V$ implies $(u^{-1}(v(x_i)), x_i) \in U$, which proves that on H, the uniform structure $\mathcal{U}_s$ is finer than the left uniform structure of H, and therefore that both structures are identical.

When H is equicontinuous, but not uniformly equicontinuous, the left uniform structure of H may be different from the structure $\mathcal{U}_s$ on H. An example is given by the following group: let u be the homeomorphism of the

real-number set $\boldsymbol{R}$, defined by $u(x) = x + 1$ if $x \leq 0$, $u(x) = x + 1/(1 + x)$ if $x \geqq 0$; and let H be the infinite cyclic group generated by u (that is, the group of all u^n, for any rational integer n). For a given $x > 0$, the sequence $(u^n(x))$ converges to $+\infty$ when n tends to $+\infty$, for it is increasing and if it had a finite limit, that limit α would satisfy the equation $\alpha = \alpha + 1/(1 + \alpha)$, which is absurd; in the same way, one sees that $u^{-n}(x)$ tends to $-\infty$ when n tends to $+\infty$, and moreover that from a given n (which depends on x) on, one has $u^{-(n+1)}(x) = u^{-n}(x) - 1$. From those remarks, one deduces easily the equicontinuity of H, and that the topology $\mathcal{T}_s$ is *discrete* on H; but $\mathcal{U}_s$ on H is not the discrete uniform structure, since $u^{n+1}(x) - u^n(x) = 1/(1 + u^n(x))$ converges to 0 when n tend to $+\infty$, for any given x.

As I have shown in a previous note [7], equicontinuity, and even uniform equicontinuity of H, is not enough to ensure that H may be completed for its left (or right) uniform structure. In the example given in that note, E is a discrete space. On the other hand, a result of R. Arens [4, p. 604], which we shall meet again in Section **11**, shows that when E is *locally compact* and *connected*, any uniformly equicontinuous group of homeomorphisms H is a dense subgroup of a complete group of homeomorphisms H' (for the left and right uniform structures). Now the example we have referred to shows that in Arens's result the assumption of connectedness may not be dropped without impairing the validity of the theorem; neither can the assumption of local compacity, as we shall now see by a slight modification of the same example. Take as E a Hilbert space, with an enumerable orthonormal basis (e_n), and let H be the group of unitary transformations of E, which permute the e_i among themselves; clearly H is uniformly equicontinuous (as any group of isometries in a metric space). Now, if u_n is the element of H such that $u_n(e_k) = e_{k+1}$ for $1 \leq k \leq n$, $u_n(e_n) = e_1$, $u_n(e_k) = e_k$ for $k > n$, Proposition 10 shows that (u_n) is a Cauchy sequence for the left uniform structure of H; but (u_n^{-1}) is not a Cauchy sequence for that structure, since $u_n^{-1}(e_1) = e_n$ has no limit in E; therefore the group H may not be completed for its left (or right) uniform structure.

Here again, consideration of the two-sided uniform structure (see Section **6**) yields less pathological results:

PROPOSITION 11. *Let H be a uniformly equicontinuous group of homeomorphisms of E, made into a topological group by the topology $\mathcal{T}_s$ on H. If E is complete, the Cauchy filters on H for the two-sided uniform structure on H, converge in the space $\mathcal{F}_s(E)$, and the set H' of their limit points is a uniformly equicontinuous group of homeomorphisms of E, which* (when topologized by $\mathcal{T}_s$) *is complete for its two-sided uniform structure, and in which H is dense.*

Indeed, let ϕ be a Cauchy filter for the two-sided uniform structure on H. For every $U \varepsilon \mathfrak{A}$, there exists a $V \varepsilon \mathfrak{A}$ such that $(x, y) \varepsilon V$ implies $(u(x), u(y)) \varepsilon U$ for every $u \varepsilon H$. Now let $(x_i)_{1 \leq i \leq n}$ be an arbitrary finite set of points of E; there exists a set $M \varepsilon \phi$ such that, for every couple (u, v) of elements of M, one has, for every i, $(u^{-1}(v(x_i)), x_i) \varepsilon V$ and $(v(u^{-1}(x_i)), x_i) \varepsilon V$; therefore, for every i, one has also $(u(x_i), v(x_i)) \varepsilon U$ and $(u^{-1}(x_i), v^{-1}(x_i)) \varepsilon U$; this proves that ϕ and the filter ϕ^{-1}, image of ϕ by $u \rightarrow u^{-1}$, are both Cauchy filters for the uniform structure $\mathcal{U}_s$. E being complete, those filters converge respectively to elements u_0, u'_0 of $\mathcal{F}_s(E)$, which belong to the closure $\bar{H}$ of H in $\mathcal{F}_s(E)$; now Lemma 5 shows that $\bar{H}$ is uniformly equicontinuous in E; therefore, the first part of the argument in Theorem 4 proves that uu^{-1} and $u^{-1}u$ converge respectively to $u_0u'_0$ and u'_0u_0 according to the filter ϕ. In the same way as in Proposition 7, this proves that u_0 is a homeomorphism of E, and $u'_0 = u_0^{-1}$. Finally, if u_0 and v_0 are limits of Cauchy filters Φ, Ψ on H, the image by the application $(u, v) \rightarrow uv$ of the product filter $\Phi \times \Psi$ is a base of a Cauchy filter on H [Bourbaki 2, p. 28] and therefore converges in the space $\mathcal{F}_s(E)$ to an element of H'; the first part of the argument of Theorem 4 shows that this element is u_0v_0, and therefore H' is a group of homeomorphisms. Now it is clear that H is dense in H', and that the two-sided uniform structure of H' gives on H the two-sided uniform structure of the group H; a classical lemma in the theory of complete uniform spaces [Bourbaki 1, p. 103] proves then that H' is complete for its two-sided uniform structure.

10. Locally compact groups of homeomorphisms. We now enlarge a little our field of investigation by considering, instead of groups of homeomorphisms of E, groups of permutations of E which are only supposed to be *continuous on every compact subset of E*, but not necessarily homeomorphisms of E^2 (i. e., subgroups contained in the space $\mathcal{P}_c(E)$). We shall give a criterion for such a group to be *locally compact*, which generalizes similar criteria concerning *compact* groups, given previously by S. B. Myers [9] and the author [8].

THEOREM 5. *Let G be a group of permutations of a uniform space E, contained in $\mathcal{P}_c(E)$* (that is, made of permutations continuous on every compact subset of E), *and such that there exists, for the topology $\mathcal{T}_c$ on G, a symmetrical neighborhood H of the identity e of G which is relatively compact in $\mathcal{P}_c(E)$. Then:*

a) *The topologies $\mathcal{T}_s$ and $\mathcal{T}_c$ are identical on G, and compatible with the group G.*

b) *G being thus topologized, the Cauchy filters on G for the right and for the left uniform structure of G are identical; each of these filters converges in the space $\mathfrak{P}_c(E)$; the set G' of limit points of these filters is a group of permutations of E* (continuous on every compact subset); *the topologies $\mathfrak{T}_s$ and $\mathfrak{T}_c$ are identical on G', and compatible with the group G'; finally, for that topology, G' is a locally compact group, the closure $\bar{H}$ of H in the space $\mathfrak{P}_c(E)$* (identical to its closure in $\mathfrak{F}_s(E)$) *being a compact neighborhood of the identity in G'; and G is a dense subgroup of G'.*

The proof of that theorem is a rather lengthy one, and we divide it in several parts.

1° We first prove that the application $(u, v) \rightarrow uv$ of $G \times G$ into itself is continuous at every point (u_0, v_0) of $G \times G$ when G is given the topology $\mathfrak{T}_c$. By assumption, there exists a $V_0 \in \mathfrak{A}$ and a compact set K_0 such that for $u \in G$ the relation $(u(x), x) \in V_0$ for every $x \in K_0$ implies $u \in H$. Now let $U \in \mathfrak{A}$ and the compact set K be arbitrarily given; let $V \in \mathfrak{A}$ be such that $V^2 \subset U$. Consider the compact subset $L = K_0 \cup v_0(K)$; by Lemma 7, the set $M = H(L)$ is relatively compact, and therefore u_0 is uniformly continuous in it; this shows that there exists a $W \subset V_0$ in $\mathfrak{A}$ such that for any couple (x, y) of elements of M such that $(x, y) \in W$, one has $(u_0(x), u_0(y)) \in V$. Now if $v \in G$ is such that $(v(x), v_0(x)) \in W$ for every $x \in K \cup v_0^{-1}(K_0)$, one has $(v(v_0^{-1}(y)), y) \in W \subset V_0$ for every $y \in K_0$ and therefore $vv_0^{-1} \in H$, from which one deduces that $v(K) \subset M$ and $v_0(K) \subset M$; therefore, one has $(u_0(v(x)), u_0(v_0(x))) \in V$ for every $x \in K$. On the other hand, let $u \in G$ be such that $(u(z), u_0(z)) \in V$ for every $z \in M$; as $v(K) \subset M$, one has $(u(v(x)), u_0(v(x))) \in V$ for every $x \in K$, and therefore $(u(v(x)), u_0(v_0(x))) \in V^2 \subset U$ for every $x \in K$, which proves the continuity of uv.

2° In the second place, we show that $u \rightarrow u^{-1}$ is continuous in G. Let K_0, V_0 have the same meaning as in 1°, $U \in \mathfrak{A}$ and the compact set K be arbitrary. Let u_0 be any element of G; consider the compact set $L = u_0^{-1}(K \cup K_0)$; the set $M = H(K \cup K_0)$ and the set $N = H(M)$ are relatively compact (Lemma 7); u_0^{-1} is uniformly continuous on N, and therefore there exists $V \in \mathfrak{A}$ such that for every couple (x, y) of points of N such that $(x, y) \in V$, one has $(u_0^{-1}(x), u_0^{-1}(y)) \in U$. On the other hand, H is uniformly equicontinuous on M, and therefore there exists $W \in \mathfrak{A}$ such that $W \subset V_0$ and that, for every couple (x, y) of points of M such that $(x, y) \in W$, and every $w \in H$, one has $(w(x), w(y)) \in V$. Now suppose $u \in G$ is such that $(u(x), u_0(x)) \in W$ for every $x \in L$; that relation may be written $(u(u_0^{-1}(y)), y) \in W$ for $y \in u_0(L)$, and as $K_0 \subset u_0(L)$, $uu_0^{-1} \in H$; this proves

first that for $x \in L$, $u(x) = u(u_0^{-1}(u_0(x)))$ and $u_0(x)$ belongs to $H(u_0(L)) = M$. On the other hand, as H is symmetrical, $u_0 u^{-1} = (u u_0^{-1})^{-1} = w$ belongs to H; therefore, as $(u(x), u_0(x)) \in W$, one has $(w(u(x)), w(u_0(x))) \in V$ for every $x \in L$; moreover, $w(u(x))$ and $w(u_0(x))$ are in $H(M) = N$, so that $(u_0^{-1}(w(u(x))), u_0^{-1}(w(u_0(x)))) \in U$ for $x \in L$; but, by definition of w, this means that $(x, u^{-1}(u_0(x))) \in U$ for $x \in L$; as this is equivalent to saying that $(u_0^{-1}(z), u^{-1}(z)) \in U$ for $z \in u_0(L)$, and in particular for $z \in K$, we have shown that $u \to u^{-1}$ is continuous at the point u_0.

3° Having seen that the topology $\mathcal{T}_c$ is compatible with the group G, it is now an easy matter to prove that, on G, $\mathcal{T}_c$ and $\mathcal{T}_s$ coincide. Indeed, as as $\mathcal{T}_c$ is finer than $\mathcal{T}_s$ on $\mathcal{F}(E)$, it will be enough to show that, on G, $\mathcal{T}_s$ is finer than $\mathcal{T}_c$. Now, for the topology $\mathcal{T}_c$ on G, a fundamental system of neighborhoods of u_0 is composed of the sets $V \cdot u_0$, where V runs through a set of neighborhoods of e in H. But, on H, the topologies $\mathcal{T}_c$ and $\mathcal{T}_s$ are identical (Lemma 6); therefore, we may suppose that a set V is composed of the $v \in H$ such that $(v(x_i), x_i) \in U$ for $U \in \mathfrak{A}$ and a finite number of points x_i; the set $V \cdot u_0$ is then composed of the $u \in G$ such that $(u(u_0^{-1}(x_i)), x_i) \in U$, which, by putting $y_i = u_0^{-1}(x_i)$, is equivalent to $(u(y_i), u_0(y_i)) \in U$; therefore every neighborhood of u_0 in G for $\mathcal{T}_c$ is also a neighborhood of u_0 in G for $\mathcal{T}_s$, which proves our contention.

We now investigate Cauchy filters on G for the left and right uniform structure of that group, and we begin by considering those structures *on* H.

4° We are going to show that, on H, the left and right uniform structure of G are *identical* to the structure $\mathcal{U}_c$ (and also to $\mathcal{U}_s$ according to Lemma 6). First, if $U \in \mathfrak{A}$ and K compact are arbitrary, $L = H(K)$ is relatively compact, and therefore there exists a $V \in \mathfrak{A}$ such that the relation $(v(x), u(x)) \in V$ for $x \in K$ implies $(u^{-1}(v(x)), x) \in U$ by the uniform equicontinuity of H on L; in the same way one sees that there exists $W \in \mathfrak{A}$ such that $(u^{-1}(v(x)), x) \in W$ for $x \in K$ implies $(v(x), u(x)) \in U$, owing to the fact that $H(L)$ is relatively compact; therefore, the structure $\mathcal{U}_c$ is identical to the left uniform structure on H. On the other hand, $(v(y), u(y)) \in U$ for $y \in L$ implies in particular $(v(u^{-1}(x)), x) \in U$ for $x \in K$; conversely, $(v(u^{-1}(y)), y) \in U$ for $y \in L$ implies in particular $(v(x), u(x)) \in U$ for $x \in K$; therefore $\mathcal{U}_c$ is also identical with the right uniform structure on H. As $\bar{H}$ is the closure of H in the space $\mathcal{P}_c(E)$, every Cauchy filter for the left (or right) uniform structure on H converges in the compact space $\bar{H}$, and $\bar{H}$ is the set of the limit points of all these filters.

5° Next we prove that $\bar{H}$ consists of permutations of E (continuous on every compact subset). We show first that $(u, v) \to uv$ is a continuous

application of $\bar{H} \times \bar{H}$ (for the topology $\mathcal{T}_c$) in the space $\mathcal{P}_c(E)$. The argument is very similar to that in 1°; let u_0, v_0 be two elements of $\bar{H}$, and let $U \varepsilon \mathfrak{A}$ and the compact subset K of E be arbitrary; $L = \bar{H}(K)$ is relatively compact, and $\bar{H}$ is therefore uniformly equicontinuous on L. Take $V \varepsilon \mathfrak{A}$ such that $V^2 \subset V$, and $W \varepsilon \mathfrak{A}$ such that for every couple (x, y) of points of L such that $(x, y) \varepsilon W$, and every $u \varepsilon \bar{H}$, $(u(x), u(y)) \varepsilon V$. Then take $v \varepsilon \bar{H}$ such that $(v(x), v_0(x)) \varepsilon W$ in K; as $v(x)$ and $v_0(x)$ are in L, one has $(u(v(x)), u(v_0(x))) \varepsilon V$ for every $u \varepsilon \bar{H}$; take $u \varepsilon \bar{H}$ such that $(u(y), u_0(y))$ εV for $y \varepsilon L$; in particular, one has $(u(v_0(x)), u_0(v_0(x))) \varepsilon V$ for every $x \varepsilon K$, and therefore $(u(v(x)), u_0(v_0(x))) \varepsilon V^2 \subset U$ for $x \varepsilon K$.

We may now apply a familiar argument already used several times. If a Cauchy filter ϕ for the left uniform structure on H converges to $u_0 \varepsilon \bar{H}$, the filter ϕ^{-1}, which is a Cauchy filter for the right uniform structure, converges to a $u'_0 \varepsilon \bar{H}$; when u tends to u_0 according to ϕ, u^{-1} tends to u'_0, and therefore, by what has just been proved, uu^{-1} and $u^{-1}u$ tend to $u_0u'_0$ and u'_0u_0 respectively; this shows that $u_0u'_0$ and u'_0u_0 are the identical application of E on itself, and therefore that u_0 and u'_0 are reciprocal permutations of E.

6° Let us now consider, more generally, Cauchy filters on G for the left or right uniform structure on that group. First let ϕ be a Cauchy filter on G for the right uniform structure; with the same meaning as in 1° for V_0 and K_0, there exists a set $M \varepsilon \phi$ such that $(u(v^{-1}(x)), x) \varepsilon V_0$ for every $x \varepsilon K_0$ and for every couple (u, v) of elements of M; this implies $uv^{-1} \varepsilon H$. Therefore, if u_0 is any element of M, one may write $\phi = \phi' u_0$, where ϕ' is a Cauchy filter for the right uniform structure, having a base made of sets of H, and therefore converging to an element w_0 of $\bar{H}$ (for the topology $\mathcal{T}_c$); one then sees immediately that ϕ converges for the same topology towards w_0u_0, which is of course a permutation of E, continuous on every compact set. Next, we show that if Ψ is a Cauchy filter on G for the left uniform structure, it is also a Cauchy filter for the right uniform structure. One sees at once as above that $\Psi = u_0\Psi'$ where Ψ' is a Cauchy filter for the left uniform structure, having a base on H, and $u_0 \varepsilon G$. Now Ψ' is also a Cauchy filter for the right uniform structure (by 4°), and therefore so is Ψ, since $v \to u_0v$ is uniformly continuous on G for the right uniform structure. It is easily verified that if Ψ' converges to $w_0 \varepsilon \bar{H}$, $\Psi = u_0\Psi'$ converges to u_0w_0 ($\bar{H}(K)$ being relatively compact for every compact subset K of E, which implies that u_0 is uniformly continuous on $\bar{H}(K)$).

Conversely, if ϕ' is any Cauchy filter on H, it is clear that $u_0\phi'$ and $\phi' u_0$ are bases of Cauchy filters on G for the left (or right) uniform structure; therefore the set G' of limit points of all Cauchy filters on G may be written $G\bar{H}$ or $\bar{H}G$. We have $G'^{-1} = (G\bar{H})^{-1} = (\bar{H})^{-1}G^{-1} = \bar{H}G = G'$. On the

other hand, if U is a neighborhood of e in H such that $U \cdot U \subset H$, one has $\bar{U} \cdot \bar{U} \subset \bar{H}$; the same argument as above shows that $G' = G\bar{U} = \bar{U}G$; therefore $G' \cdot G' = G(\bar{U}G)\bar{U} = (G \cdot G)(\bar{U} \cdot \bar{U}) \subset G \cdot \bar{H} = G'$, which shows that G' is a *group* of permutations of E (continuous on every compact set). It is obvious that G is dense in G' (for the topology $\mathcal{T}_c$), and that $\bar{H}$ is a symmetrical neighborhood of e in G' for that topology; the three first parts of the proof, applied to G' instead of G, show at once that on G' the topologies $\mathcal{T}_c$ and $\mathcal{T}_s$ are identical and compatible with G'; of course, $\bar{H}$ being compact (by Lemma 7) G' is a locally compact group. The proof of Theorem 5 is thus completed.

The group G' is of course contained in the closure $\bar{G}$ of G in the space $\mathcal{P}_c(E)$; but it must be stressed that in general G' is not identical to $\bar{G}$, and that accounts for the rather devious way in which we have been compelled to define G'. For instance, the general linear group on Cartesian n-space E is locally compact for the topology $\mathcal{T}_c$, but its closure in $\mathcal{B}_c(E)$ *is not a group of homeomorphisms*; in fact, it is easy to see that it is the ring of all linear applications of E in itself. This proves also that on G the uniform structure $\mathcal{U}_c$ (as well as $\mathcal{U}_s$) are *distinct* from both left and right uniform structures of G; this may of course be seen directly on the above example: in fact, when E is one-dimensional, the left and right uniform structure are identical with the multiplicative uniform structure on the (multiplicative) group of real numbers $\neq 0$, whilst $\mathcal{U}_s$ and $\mathcal{U}_c$ are identical with the additive uniform structure on that set of real numbers, and it is well known that those two structures are not even comparable [Bourbaki 2, p. 54].

We observe, finally, that in Theorem 5 the assumption that H is *symmetrical* (that is, $H^{-1} = H$) cannot be dispensed with; it is indeed easy to verify that in the group G of homeomorphisms generated by a set of homeomorphisms given as an example by R. Arens [4, p. 601], there exists a non-symmetrical compact neighborhood of the identity; but in G, u^{-1} is not even continuous.

11. Locally compact groups of homeomorphisms of locally compact and connected spaces. We shall now see, generalizing results of van der Waerden and van Dantzig [10], R. Arens [4, p. 604] and S. B. Myers [9, p. 498-499], that it is possible to relax the assumptions on G in Theorem 5, when one imposes on the space E the restrictions that it be *locally compact* and *connected.*

PROPOSITION 12. *Let E be a locally compact and connected space, G a group of homeomorphisms of E. If there exists, for the topology $\mathcal{T}_c$, a symmetrical neighborhood H_0 of the identity e of G, and a uniform structure $\mathcal{U}$ on E such that one of the following conditions is satisfied:*

a) *E is complete and H_0 equicontinuous in E for $\mathcal{U}$;*

b) *H_0 is uniformly equicontinuous in E for $\mathcal{U}$;*

then G satisfies the assumptions of Theorem 5.

We proceed to show that there exists a symmetrical neighborhood $H \subset H_0$ of e in G (for the topology $\mathcal{T}_c$) which is relatively compact in $\mathcal{C}_c(E)$. There exists a $V_0 \in \mathfrak{A}$ and a compact set K_0 in E such that the set L of $u \in G$ for which $(u(x), x) \in V_0$ for all $x \in K_0$ is contained in H_0; moreover, we may suppose that V_0 has been taken such that $V_0(K_0)$ is relatively compact in E (if not, replace V_0 by another set of $\mathfrak{A}$ contained in V_0 and satisfying the preceding condition). The set H_0 being equicontinuous on $V_0(K_0)$, is uniformly equicontinuous on that relatively compact set; therefore, as H_0 is symmetrical, there exists a $W \in \mathfrak{A}$ such that for all $u \in L$ satisfying $(u(x), x) \in W$ for all $x \in K_0$, one has $(x, u^{-1}(x)) \in V_0$ for all $x \in K_0$; this means of course that L contains a symmetrical neighborhood H of e in G such that, for all $x \in K_0$, $H(x)$ is a relatively compact set in E. It remains to be proved that for *all* $x \in E$, $H(x)$ is relatively compact. We consider for that purpose the subset A of the points x in E such that $H(x)$ is relatively compact; we wish to prove that in both cases considered in Proposition 12, A is open and closed in E; as E is connected and $K_0 \subset A$ (therefore A is not the empty set), this will show that $A = E$, and therefore complete the proof.

We first proceed to prove that A is *open* under the only assumption that H is *equicontinuous* in E. Let x_0 be any point of A, and let $M = H(x_0)$; $\bar{M}$ is compact, and therefore, as E is locally compact, there exists a $V \in \mathfrak{A}$ such that $V(\bar{M})$ is relatively compact [Bourbaki 1, p. 111]; H being equicontinuous at x_0, there exists a neighborhood U of x_0 such that, for $x \in U$ and for every $u \in H$, one has $(u(x), u(x_0)) \in V$, and therefore $H(x) \subset V(\bar{M})$; therefore $H(x)$ is relatively compact for every $x \in U$.

The proof that A is closed is not so easy, and we have to consider separately assumptions a) and b).

a) Suppose that x_0 belongs to the closure of A; let ϕ be an ultrafilter on H; $\phi(x_0)$ is then an ultrafilter on $H(x_0)$, and conversely every ultrafilter on $H(x_0)$ may be obtained in that way [Bourbaki 1, p. 28]; therefore, to show that $H(x_0)$ is relatively compact, it is enough to prove that $\phi(x_0)$ converges in E [Bourbaki 1, p. 59], and as E is complete, this is equivalent to proving that $\phi(x_0)$ is a Cauchy filter base on E. Now, H being equicontinuous at x_0, for every $V \in \mathfrak{A}$, there exists a neighborhood U of x_0 such that, for every $x \in U$ and every $u \in H$, one has $(u(x), u(x_0)) \in V$; there exists an $x \in A$ belonging to U, and therefore $H(x)$ is relatively compact; from this one deduces that the ultrafilter base $\phi(x)$ converges to a point $z \in E$. There

exists, therefore, a set $M \varepsilon \phi$ such that for every couple (u, v) of elements of M, one has $(u(x), v(x)) \varepsilon V$; one has also $(u(x_0), v(x_0)) \varepsilon V^3$, which proves that $\phi(x_0)$ is a Cauchy filter base on E.

b) Let x_0 and ϕ have the same meaning as in a), and $W \varepsilon \mathfrak{A}$ be such that $W(x_0)$ is relatively compact; take $V \varepsilon \mathfrak{A}$ such that the relation $(y, z) \varepsilon V^3$ implies $(u(y), u(z)) \varepsilon W$ for every $u \varepsilon H$ (uniform equicontinuity of H). The same argument as in a) shows that there exists a set $M \varepsilon \phi$ such that, for every couple (u, v) of elements of M, one has $(u(x_0), v(x_0)) \varepsilon V^3$; take any $u_0 \varepsilon M$; one has $(x_0, u_0^{-1}(v(x_0))) \varepsilon W$ for every $v \varepsilon M$, or $u_0^{-1}(v(x_0)) \varepsilon W(x_0)$; therefore for every $v \varepsilon M$, $v(x_0)$ belongs to the relatively compact set $u_0(W(x_0))$, and therefore $\phi(x_0)$ converges in E.

One is naturally led to ask if Proposition 12 is still true when it is only supposed that H_0 is *equicontinuous* in E (E being locally compact and connected). We shall now give an example showing that the answer to that question is *negative.*

We first define a locally compact and connected subspace F of ordinary 3-space in the following way: let C_n be the closed disc: $z = 1/n$, $x^2 + (y - n)^2 \leq n^2$ for any integer $n \geq 2$; F is the union of the C_n for every $n \geq 2$, of the closed half-plane $z = 0$, $y \geq 0$, and of the segments joining the point S: $x = 0$, $y = 0$, $z = 1$ to the center of C_2, and the center of each C_n to the center of C_{n+1} for every $n \geq 2$. For every integer p, we define an homeomorphism u_p of F as follows: u_p leaves invariant S and the segments joining S and the centers of the C_n; in C_n, u_p is a rotation about the centre of C_n, of an angle p/n; finally, in the half-plane $z = 0$, $y \geq 0$, u_p is the translation of vector p parallel to the x axis; it is easily seen that u_p is continuous in F and that u_{-p} is the reciprocal of u_p; moreover, it is clear that the u_p form a group algebraically isomorphic to the additive group of integers.

Now, for each finite sequence of rational integers $(n_i)_{1 \leq i \leq k}$ (k arbitrary), consider a "copy" $F_{n_1 n_2 \ldots n_k}$ of the space F; for any two of those spaces, there is a canonical homeomorphism of one on the other, which to every point of one associates the point in the other having the same coordinates. Out of the "topological sum" [Bourbaki 1, p. 50] of the enumerable set formed of F and the spaces $F_{n_1 n_2 \ldots n_k}$, we make a connected and locally compact space E by the following procedure of "fastening" them together: we identify the "summit" S of the space $F_{n_1 \ldots n_k n_{k+1}}$ with the point of coordinates $(n_{k+1}, 0, 0)$ in $F_{n_1 \ldots n_k}$, and the summit of F_n with the point $(n, 0, 0)$ in F. Next, we define homeomorphisms $u_{n_1 \ldots n_k, p}$ of E in the following way: $u_{n_1 \ldots n_k, p}$ leaves invariant each point of every $F_{m_1 \ldots m_q}$ for which the sequence of the m's has less than k terms, or has at least k terms but has its first k terms distinct

from $n_1, n_2, \cdots, n_k$; on $F_{n_1 n_2 \ldots n_k}$, $u_{n_1 n_2 \ldots n_k, p}$ operates as does u_p on F; finally, on every $F_{n_1 \ldots n_k n_{k+1} n_{k+2} \ldots n_r}$, $u_{n_1 \ldots n_k, p}$ is identical with the canonical homeomorphism of that space onto $F_{n_1, \ldots, n_k, n_{k+1}+p, \ldots, n_r}$. Let G be the group of homeomorphisms of E generated by the u_p and $u_{n_1 \ldots n_k, p}$; it is clear that the effect of any homeomorphism of the group G on one of the F's is to send it (canonically) into another of the F's (eventually the same) and then to operate on that F as one of the u_p.

We now put on E the uniform structure induced on it by the (unique) uniform structure of the compact space E^* obtained by the adjunction to E of a "point at infinity." One sees very easily that, for that uniform structure, the set of the u_p is equicontinuous at every point of F (for any point x in the half-plane $z = 0$, $y \geq 0$, any compact subset K in F, and any compact neighborhood U of x in F, all the sets $u_p(U)$ except a finite number are outside K); from this it follows immediately that G is also equicontinuous (taking into account that a compact subset of E may intersect only a finite number of the $F_{n_1 n_2 \ldots n_k}$). But there is no neighborhood H of the identity e in G (for the topology $\mathcal{T}_c$) such that $H(x)$ is relatively compact for every $x \in E$; for a fundamental system of neighborhoods of e is formed by the subsets of G leaving invariant the points of a finite number of the $F_{n_1 n_2 \ldots n_k}$; let $F_{n_1 n_2 \ldots n_k}$ be one of those subspaces having the maximum number of indices: then all the $u_{n_1 n_2 \ldots n_k n_{k+1}, p}$ belong to the neighborhood considered, and they send the point $(0, 0, 0)$ in $F_{n_1 \ldots n_{k+1}}$ into a set of points which is not relatively compact.

In the special case where G itself is an *equicontinuous* group of homeomorphisms of the (locally compact and connected) space E, and the conditions of Proposition 12 are satisfied, the locally compact group G' defined in Theorem 5 is identical with the *closure* $\bar{G}$ of G in the space $\mathcal{C}_c(E)$ (or the space $\mathcal{F}_s(E)$); this is a generalization of a result of R. Arens [4, p. 604], who considers only the case when G is uniformly equicontinuous in E; the property then results immediately from Proposition 10 and Theorem 5, since the Cauchy filters for $\mathcal{U}_s$ and for the left uniform structure on G are then identical. In the case when G is only supposed to be equicontinuous in E, we have seen that in general the left uniform structure on G is strictly finer than $\mathcal{U}_s$ (Section **9**); therefore every Cauchy filter on G for the left uniform structure is a Cauchy filter for $\mathcal{U}_s$. Conversely, we shall prove that every *convergent* (in $\mathcal{F}_s(E)$) Cauchy filter for $\mathcal{U}_s$ on G is also a Cauchy filter for the left uniform structure; as we know from Theorem 5 that G' is the set of limit points of all Cauchy filters for the left uniform structure, it will follow that $G' = \bar{G}$. Let, therefore, ϕ be a Cauchy filter for $\mathcal{U}_s$ on G, and let it converge to $u_0 \in \mathcal{F}_s(E)$; for every finite set (x_i) of points of E, each of the filter bases $\phi(x_i)$ converges in E to $u_0(x_i) = y_i$; for every $V \in \mathfrak{A}$, there exists

therefore a set $M \varepsilon \phi$ such that, for every $u \varepsilon M$ and every i, one has $(u(x_i), y_i) \varepsilon V$. Now let U be an arbitrary set of $\mathfrak{A}$, and $W \varepsilon \mathfrak{A}$ be such that $W^2 \subset U$; as G is equicontinuous in E, there exists a $V \varepsilon \mathfrak{A}$ such that, for $(z, y_i) \varepsilon V$, one has $(u(z), u(y_i)) \varepsilon W$ for every i and every $u \varepsilon G$; if M is a set in ϕ corresponding to V, we see thus that for every couple (u, v) of elements of M, one has $(x_i, u^{-1}(y_i)) \varepsilon W$ and $(u^{-1}(v(x_i)), u^{-1}(y_i)) \varepsilon W$, and therefore $(u^{-1}(v(x_i)), x_i) \varepsilon W^2 \subset U$ for every i, which shows that ϕ is a Cauchy filter for the left uniform structure on G (taking into account that, on G, by Lemma 6, $\mathcal{U}_s$ and $\mathcal{U}_c$ are identical).

We conclude with some comments on the relation, on a Lie group of transformations G of an analytic manifold E, between the topology $\mathcal{T}$ defined on G by the parameters, and the topology $\mathcal{T}_c$ (see R. Arens [4, p. 608]). By assumption, there exists for the topology $\mathcal{T}$, a compact symmetrical neighborhood U of the identity e in G, such that, for any point $x_0 \varepsilon E$, there exists a neighborhood V of x_0 in E, U and V being both homeomorphic with compact neighborhoods of cartesian spaces, the application $(x, u) \to u(x)$ of $V \times U$ into E carrying $V \times U$ into a subset of another neighborhood W of x_0 (also homeomorphic to a cartesian neighborhood), and being moreover analytic for a suitable system of local cordinates in U, V and W. The continuity of $(u, x) \to u(x)$ in $U \times E$ shows that, *on the set* U, the topology $\mathcal{T}$ is *finer* than $\mathcal{T}_c$ [4, p. 596]; as U is compact for $\mathcal{T}$, $\mathcal{T}$ and $\mathcal{T}_c$ are identical on U. From that, it follows that a necessary and sufficient condition for $\mathcal{T}$ and $\mathcal{T}_c$ to be *identical on* G, is that U must be *a neighborhood of* e *in* G *for the topology* $\mathcal{T}_c$. The necessity of the condition is obvious; conversely, if it is verified, G satisfies the conditions of Theorem 5, and moreover is identical with the group G' defined in that theorem, since U is, by assumption, compact and therefore closed in $\mathcal{B}_c(E)$; this shows that $\mathcal{T}_c$ is compatible with the group G, and gives the same set of neighborhoods of e as $\mathcal{T}$, and therefore is identical with $\mathcal{T}$.

The preceding condition may be worded as follows: there must exist a compact subset K of E, and a $T \varepsilon \mathfrak{A}$ such that the set of homeomorphisms $u \varepsilon G$ having the property that $(u(x), x) \varepsilon T$ for every $x \varepsilon K$ is contained in U; one can say roughly that only homeomorphisms of the group having parameters sufficiently near to those of the identity, may displace a little all points of any compact subset. Of course, that property is not possessed by all Lie groups of transformations. For instance, if E is a two-dimensional torus, considered as the set of all points (x, y) of the plane, where x and y are taken modulo 1, the group G of the homeomorphisms which send (x, y) into $(x + t, y + \theta t)$ taken mod. 1 (t being any real number, θ a fixed real number) is a one-parameter Lie group, which for $\mathcal{T}$ is isomorphic to the

topological additive group of all real numbers. But, if θ is taken irrational, there exist values of t as large as we please such that both t and θt, taken mod. 1, are as small as we please. Therefore $\mathcal{J}$ and $\mathcal{J}_c$ are distinct on G (one may see easily that for $\mathcal{J}_c$, G is isomorphic to a dense subgroup of the two-dimensional torus group).

UNIVERSITY OF SÃO PAULO,
SAO PAULO (BRAZIL).

BIBLIOGRAPHY.

[1] N. Bourbaki, *Eléments de Mathématique, Topologie générale,* Chap. I-II (Paris, Hermann, 1940).

[2] N. Bourbaki, *Eléments de Mathématique, Topologie générale,* Chap. III-IV (Paris, Hermann, 1942).

[3] R. Arens, "A topology for spaces of transformations," *Annals of Mathematics,* vol. 47 (1946), pp. 480-495.

[4] R. Arens, "Topologies for homeomorphism groups," *American Journal of Mathematics,* vol. 68 (1946), pp. 593-610.

[5] G. Birkhoff, "The topology of transformation sets," *Annals of Mathematics,* vol. 35 (1934), pp. 861-875.

[6] J. Braconnier and J. Colmez, "Sur les groupes d'homéomorphismes d'un espace complètement régulier," *Comptes-rendus des séances de l'Académie des Sciences de Paris,* vol. 223 (1946), pp. 230-232.

[7] J. Dieudonné, "Sur la complétion des groupes topologiques," *Comptes-rendus des séances de l'Académie des Sciences de Paris,* vol. 218 (1944), pp. 774-776.

[8] J. Dieudonné, "Sur les groupes compacts d'homéomorphismes," *Anais da Academia Brasileira de Ciências,* vol. 18 (1946), pp. 287-289.

[9] S. B. Myers, "Equicontinuous sets of mappings," *Annals of Mathematics,* vol. 47 (1946), pp. 496-502.

[10] B. L. van der Waerden and D. van Dantzig, "Über metrisch homogene Raüme," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität,* vol. 6 (1928), pp. 367-376.

[11] A. Weil, "Sur les espaces à structure uniforme," *Actualités Scientifiques et Industrielles,* no. 551, Paris, Hermann, 1937.

[12] A. Weil, "L'intégration dans les groupes topologiques et ses applications," *Actualités Scientifiques et Industrielles,* no. 869, Paris, Hermann, 1940.

ERRATA TO THE PAPER OF L. FEJES TÓTH.

p. 177, 5 line from below. After c_i insert "lying outside ϕ_i."

p. 177, 2 line from below. In the last integral instead of ϕ_i write "ϕ'_i."

p. 177, last line. Instead of "where ϕ_i" write "where ϕ'_i."

ON ORDERED LOOPS.*

By DANIEL ZELINSKY.

In [6] [1] a valuation of a nonassociative algebra was defined as a function on the algebra to an ordered loop with the usual postulates on the values of sums and products. It was shown that a necessary and sufficient condition for an ordered loop L to be the value loop of some algebra with a unity quantity and of finite order over a field is the following: Some subgroup G of the center of L has finite index in L. In the present paper we determine all such ordered loops L.

The key lemma is Theorem 3 which asserts the existence of a nonzero l-ideal of L lying completely in the center of L. The proof consists merely in noting that L is a topological loop whose center is open and that the l-ideals usually form a set of neighborhoods of zero. The major result follows simply from Theorem 3 and states that L and L/G are centrally nilpotent. From here it is a matter of computation to show that every such L is the result of a special sequence of loop extensions determined by G, the (abelian) factor groups of L/G and three factor sets of specified types. Conversely, given an ordered group G, a set of finite abelian groups (almost arbitrarily chosen) and three factor sets (two of these are functions of certain specified variables, but otherwise arbitrary, and the third is of a specified form) this same sequence of loop extensions builds an ordered loop of the type required. These loops, combined with the results of [6] give a large class of nonassociative algebras with valuations.

However, an interesting corollary of this theory, derived at the end of section 3, implies that any valuation of an algebra of finite order over an algebraic number field necessarily has a value loop order-isomorphic with an additive group of real numbers.

1. Topological loops. A loop is a set L of elements $x, y, z, \cdots$, with the properties (1°) for every x and y of L there is a unique element $x+y$ of L; (2°) for every x and y of L there is a unique element u, denoted by $x-y$, such that $u+y=x$; (3°) for every x and y of L there is a unique

* Received July 8, 1947.

[1] Numbers in brackets refer to the references cited at the end of the paper.

element v, denoted by $-y+x$, such that $y+v=x$; and (4°) there is an element 0 such that $0+x=x+0=x$ for all x of L.

A topological loop is a loop that is also a T_0-space in which $x+y$, $x-y$ and $-y+x$ are continuous functions of x and y. In particular, this definition guarantees that the transformations R_x and L_x defined by

$$yR_x = y + x, \qquad yL_x = x + y$$

are homeomorphisms of the topological space L. For R_x and L_x are clearly continuous transformations and R_x^{-1} and L_x^{-1} are precisely the transformations

$$yR_x^{-1} = y - x, \qquad yL_x^{-1} = -x + y,$$

which are continuous by hypothesis. In a similar fashion the transformations $x \to y - x$ and $x \to -x + y$, for fixed y, are homeomorphisms. Hence local properties of a topological loop can be proved merely by proving them in the neighborhood of 0; specifying a set of neighborhoods of 0 completely determines the topology of a topological loop.

The ordinary proofs [5, pp. 42-43] then guarantee that every topological loop is a Hausdorff space. This is enough to prove

LEMMA 1. *A finite topological loop is discrete.*

One further theorem that we must carry over from the theory of topological groups concerns the topologization of quotient groups. If H is a closed normal subloop [2] of the topological loop L then we may impose on the quotient loop L/H the "natural topology": a set is open in L/H if and only if its inverse image is open in L.

LEMMA 2. *If H is a closed normal subloop of the topological loop L then the quotient loop L/H is a topological loop under the natural topology and the natural mapping of L onto L/H is a continuous, open homomorphism.*

The proof in the associative case remains valid verbatim [5, pp. 45-46].

LEMMA 3. *The center* [3] *of a topological loop is closed.*

[2] A subloop H of a loop L is a normal subloop or a normal divisor of L if it satisfies any one of the following three equivalent conditions: (1°) H is the kernel of a homomorphism of L; (2°) for all x, y of L, $x+(y+H)=(x+y)+H=(H+x)+y$; (3°) H is invariant under all inner mappings of L. This agrees with the definitions given by Albert [1, p. 513] and Bruck [3, p. 256]; condition (3°) is due to Bruck.

[3] The center of L is the set of all elements of L which are invariant under the inner mappings of L; equivalently, the center is the set of all c for which $c+(y+z)=(c+y)+z=y+(z+c)$ for all y and z of L. If c is in the center of L then c commutes with all elements of L and every sum of three elements with c as one of the addends is associative.

Proof. Let us call the loop L and its center C. We must show that if a is an element of the closure of C, then a itself is in C; that is, for all y and z of L and for $x = a$, the following expressions are zero: $[x + (y + z)] - [(x + y) + z]$, $[x + (y + z)] - [y + (z + x)]$. Let $q(x)$ denote either one of these expressions, thought of as a function of x with y and z fixed. The function $q(x)$ is continuous and so for every neighborhood U of $q(a)$ there is a neighborhood V of a such that $q(V) \subset U$. Since a is in the closure of C, every such V contains a point c of C and $0 = q(c)$ is in $q(V) \subset U$. Thus 0 is in every neighborhood U of $q(a)$, $q(a) = 0$, proving that a is in C.

2. **Ordered loops.** An ordered loop L is a loop in which is defined a binary relation $<$ with the properties that for all x, y, z of L, (1°) $x < y$ and $y < z$ imply $x < z$; (2°) one and only one of the statements $x < y$, $y < x$, $x = y$, is true; and (3°) $x < y$ implies $x + z < y + z$ and $z + x < z + y$. We shall use the ordinary language in connection with this relation, $<$. For example, "$x < y$" will be read "x is less than y," $x > y$ will mean $y < x$, "$x > 0$" will be expressed by "x is positive," etc. If A and B are subsets of L, $A > B$ will mean that every element of A is greater than every element of B.

Ordered loops have all the elementary properties of ordered groups (linearly ordered l-groups). In particular, the transformations R_x, L_x, R_x^{-1}, L_x^{-1} are all order preserving (lattice automorphisms) and so also are all the transformations of the "group associated with L" [3, p. 257; 1, pp. 511-512], since these latter are just finite products of R's and L's and inverses of R's and L's. We may also remark that if $x < y$ then $z - x > z - y$ and $-x + z > -y + z$ as in the case of ordered groups.

An l-ideal or isolated subloop of an ordered loop L is a normal subloop M of L with the property that if x and y belong to M and $x < z < y$, then z belongs to M. In case L is a group, this definition of an l-ideal coincides with that given by Birkhoff [2, p. 310]. The following lemma is merely the contrapositive form of our definition.

LEMMA 4. *If M is an l-ideal of L and x is an element of L that is not in M, then either $x < M$ or $x > M$.*

Much as in the case of groups [4, p. 172; 2, p. 310], a straightforward proof based on Lemma 4 shows that if M is an l-ideal of L, the quotient loop L/M can be ordered by defining the inequality $x + M < y + M$ in L/M to mean that the sets $x + M$ and $y + M$ have this relation as subsets of L.

The l-ideal M is then the kernel of an order-preserving homomorphism of L into an ordered loop; and indeed this property of M is easily seen to be equivalent to the condition that M be an l-ideal of L.

Similarly, on the basis of Lemma 4, we can show that the set of l-ideals of L is linearly ordered by inclusion. For if M_1 and M_2 are two l-ideals and M_1 is not contained in M_2, then there is an x in M_1 that is not in M_2, $x < M_2 < 0 - x$ or $x > M_2 > 0 - x$, so that M_2 is contained in M_1. A trivial conclusion from this remark is

LEMMA 5. *The set-theoretic union of a set of l-ideals of L is again an l-ideal of L.*

We also omit the proof of

LEMMA 6. *The intersection of a set of l-ideals of L is again an l-ideal of L.*

Every subset S of L can then be said to generate an l-ideal, namely, the intersection of all l-ideals containing S.

LEMMA 7. *If S is a normal subloop of L, the l-ideal generated by S is the set S' consisting of all x in L such that, for some elements y_x and z_x of S, $y_x \leqq x \leqq z_x$.*

Proof. Clearly S' is contained in the l-ideal generated by S. If S' is a normal subloop of L, the truth of the lemma is obvious. But S' is invariant under any inner mapping of L. For if J is an inner mapping and x is an element of S', then for some y and z of S, $y \leqq x \leqq z$, $yJ \leqq xJ \leqq zJ$ (since, by the second paragraph of this section, J is order-preserving), yJ and zJ are in S, so that xJ is in S'.

Again as in the case of groups, an ordered loop with more than one element is a topological loop under the interval topology. More precisely, if we use the term "open interval (y, z)" for the set of all x in L with $y < x < z$, then L becomes a topological space when we designate the set of all open intervals as an open base. For the intersection of two open intervals is an open interval.

THEOREM 1. *An ordered loop L having at least two elements is a topological loop under the interval topology described above.*

Proof. We must show that L is a T_0-space and that $x + y$, $x - y$ and $-y + x$ are continuous functions of x and y. To show that L is a T_0-space, let x and y be distinct points of L and, say, $x < y$. There is an element z

in L with $z < x$ (if $x < 0$, let $z = x + x$; if $x \geqq 0$, choose any $z < 0$; such a z exists since L is supposed to have a nonzero element z_0 and either $z_0 < 0$ or $0 - z_0 < 0$). Then the open interval (z, y) contains x but not y. In demonstrating the continuity of $x + y$, $x - y$ and $-y + x$ we distinguish two cases. First, if the set consisting of 0 alone is open then every point forms an open set (since R_x is a homeomorphism and carries 0 into $0 + x = x$) and the topology is the discrete topology. Clearly L is then a topological loop. Hence let us assume that the topology is not discrete.

LEMMA 8. *The interval topology in an ordered loop L is the discrete topology if and only if L has a least positive element, e. In this case, e is in the center of L.*

Proof. If the topology is discrete then $\{0\}$ is open and there is an interval (x, y) containing only 0. Then $e = y$ is a least positive element of L. Conversely, if y is the least positive element of L, then $0 - y$ is the largest negative element of L. For $0 - y < z < 0$ implies $-(0 - y) + 0 > -z + 0 > 0$, or $y > -z + 0 > 0$, a contradiction. Then the interval $(0 - y, y)$ contains only 0, $\{0\}$ is open and the topology is discrete.

The proof that e is in the center of L is contained in a previous paper [6], but will be reproduced here for the sake of completeness. We must prove that $e + (x + y) = (e + x) + y = x + (y + e)$ for all x, y of L. To do this, we first show that there is no element z with $x + y < z < e + (x + y)$ or $x + y < z < (e + x) + y$ or $x + y < z < x + (y + e)$. For example, if $x + y < z < x + (y + e)$, then $y < -x + z < y + e$, $0 < -y + (-x + z) < e$, which is impossible. In similar fashion we arrive at the conclusion that of the three elements $e + (x + y)$, $(e + x) + y$, $x + (y + e)$, all are greater than $x + y$ but none can be greater than any other. Hence they are equal.

Therefore, we shall use the phrase "discrete ordered loop" to mean "ordered loop with a least positive element." Note that this terminology differs from that of Krull [4, pp. 171 f.].

If we are to prove Theorem 1 for a nondiscrete loop L, we may assume that between every two elements there is a third. For if $x < z$, then $0 < z - x$ so that there is a y' between 0 and $z - x$. Then $y = y' + x$ is between x and z. The proofs of the continuity of the functions $x + y$, $x - y$ and $-y + x$ are very similar. Let us give only the last here. Suppose (u, v) is an open interval containing $-y_0 + x_0$. From $u < -y_0 + x_0 < v$, we deduce $y_0 + u < x_0 < y_0 + v$. Hence there exist u', v' such that $y_0 + u < u' < x_0 < v' < y_0 + v$. Some of these inequalities may be rewritten to

give $v' - v < y_0 < u' - u$ so that there exist u'', v'' with $v' - v < u'' < y_0 < v'' < u' - u$. We shall show that if x is in the interval (u', v') and y is in the interval (u'', v'') then $-y + x$ is in (u, v), proving the continuity of this function. For such an x and y, $-y + x$ certainly lies between $-v'' + u'$ and $-u'' + v'$ so that we need only show $u < -v'' + u' < -u'' + v' < v$. But from the definition of u'' we have $v' < u'' + v$, $-u'' + v' < v$; $-v'' + u' < -u'' + v'$ since $u' < v'$ and $u'' < v''$; and the definition of v'' shows that $v'' + u < u'$, $u < -v'' + u'$. This proves Theorem 1.

As remarked above, the topology of a topological loop is completely determined once the set of neighborhoods of 0 is given. Up to this point, our neighborhoods of 0 in an ordered loop are the intervals (x, y) containing 0. However, it will frequently be convenient to replace this set by an equivalent set $\mathfrak{M}$ of neighborhoods. A class $\mathfrak{M}$ of sets each containing 0 is called an equivalent set of neighborhoods of 0 in case every open interval containing 0 contains a member of $\mathfrak{M}$ and every member of $\mathfrak{M}$ contains an open interval containing 0.

THEOREM 2. *Let L be an ordered loop considered as a topological loop under the interval topology and denote by M^0 the intersection of all the nonzero l-ideals of L. If M^0 consists of 0 alone, then the set $\mathfrak{M}$ of nonzero l-ideals of L forms an equivalent set of neighborhoods of 0. If M^0 is not 0 then an equivalent set of neighborhoods of 0 is the class of all open intervals (x, y) with x and y in M^0.*

Proof. Let $\mathfrak{M}$ denote the class of nonzero l-ideals of L. Then every element of $\mathfrak{M}$ contains a positive and a negative element of L and hence contains an open interval about zero. Then to prove Theorem 2 when $M^0 = 0$, we need only show that every open interval about 0 contains an element of $\mathfrak{M}$. Let (x, y) be such an interval. Since $M^0 = 0$, there is an M' in $\mathfrak{M}$ that fails to contain x and an M'' in $\mathfrak{M}$ that fails to contain y. If M denotes the smaller of M' and M'', then M is in $\mathfrak{M}$ and contains neither x nor y. Then by Lemma 4, $x < M$ or $x > M$; but $x < 0$ so that necessarily $x < M$. Similarly, $y > M$ and so $M \subset (x, y)$.

When $M^0 \neq 0$, every interval (x, y) about 0 contains an interval (x', y') about 0 with x' and y' in M^0. For if x is in M^0, choose $x' = x$; if x is not in M^0, then by Lemmas 4 and 6, $x < M^0$ and we may choose x' as any negative quantity of M^0. Similarly, choose $y' = y$ or y' an arbitrary positive element of M^0, according as y is or is not in M^0. Then $(x', y') \subset (x, y)$. This proves Theorem 2.

3. Ordered loops with centers of finite index. Henceforth we consider only ordered loops such as arise as value loops of algebras of finite order, namely, ordered loops L such that the center C of L contains a group G (which is then a normal subloop of L) with L/G finite.

LEMMA 9. *The quotient loop L/C is finite.*

Proof. L/C is a homomorph of the finite loop L/G.

By Lemma 3, C is closed. By Lemma 2, L/C is a topological loop and the mapping of L onto L/C is continuous. But L/C is necessarily discrete (Lemmas 1 and 9) so that its identity element $\bar{0}$ is open. It follows that the inverse image of $\bar{0}$ in L, namely C, is also open. We have proved

LEMMA 10. *The center C of L is open.*

We are now prepared to prove a key theorem.

THEOREM 3. *If L is an ordered loop not consisting of 0 alone, with center C and with L/C finite, then there is a unique nonzero l-ideal M_1 of L, contained in C and maximal in the sense that M_1 contains every other l-ideal contained in C.*

Proof. Define M^0 as in Theorem 2. If $M^0 = 0$ then the nonzero l-ideals of L form a set of neighborhoods of 0 by Theorem 2. But C is open and contains 0 by Lemma 10, so that some M is contained in C. In case $M^0 \neq 0$ a similar argument shows that some open interval (x, y) is contained in C with both x and y in M^0. If the interval $(0, y)$ is void then y is the least positive element of L and is in the center of L by Lemma 8; in this event, let z be defined to be y. If the interval $(0, y)$ is not void, let z be any element of $(0, y)$. Whichever way z is defined, we know that it is an element of $M^0 \cap C$. The normal subloop of L generated by z is then simply the set of all integral multiples of z and by Lemma 7 the l-ideal M generated by z consists of all w in L with $mz \leqq w \leqq nz$ for some integers m and n. Since z in in M^0 and M^0 is an l-ideal (Lemma 6), surely $M \subset M^0$ and $M \neq 0$. But M^0 is the smallest nonzero l-ideal, so $M = M^0$. Then every element of M^0 is of the form $mz + w$ where m is an integer and $0 \leqq w < z$. Since w is in the interval (x, y), w is in C; but mz is also in C and so $M^0 \subset C$. This proves that in every case there is a nonzero l-ideal of L in C. Let M_1 be the union of all such l-ideals. By Lemma 5, M_1 is an l-ideal and obviously satisfies the other conditions of the theorem.

THEOREM 4. *Suppose L is an ordered loop containing a group G in its*

center with L/G finite. Then there is a unique finite chain $0 = M_0 \subset M_1 \subset M_2 \subset \cdots \subset M_n = L$ of l-ideals of L where M_i $(i = 1, 2, \cdots, n)$ is characterized as the unique set M, maximal with respect to the conditions

(I$_i$) *M is an l-ideal of L properly containing M_{i-1},*

(II$_i$) *M/M_{i-1} is in the center of L/M_{i-1};*

that is, M_i satisfies conditions (I$_i$) *and* (II$_i$) *and contains every M that satisfies* (I$_i$) *and* (II$_i$).

The loop $K = L/G$ has a like property: There exist normal subloops H_i of K with $0 = H_0 \subset H_1 \subset H_2 \subset \cdots \subset H_n = K$, all inclusions after the first being proper and such that H_i/H_{i-1} is in the center of K/H_{i-1} $(i = 1, 2, \cdots, n)$. In other words K is centrally nilpotent.[4]

Proof. By Lemma 9 and Theorem 3, there is a unique M_1 maximal with respect to (I$_1$) and (II$_1$). To define the M's inductively, assume the existence of a unique maximal M_i satisfying (I$_i$) and (II$_i$), and consider $L_i = L/M_i$. This quotient loop is again an ordered loop and if G_i denotes the image of G in L_i under the natural mapping, then it is clear that G_i is a subgroup of the center of L_i. Furthermore, L_i/G_i is a homomorph of L/G and so is finite. Then by Theorem 3, either $L_i = 0$, so that $M_i = L$, or there is a unique M'_{i+1} in the center of L_i which is a nonzero l-ideal of L_i and is maximal with respect to these properties. Define M_{i+1} to be the inverse image of M'_{i+1} in L. Then it is trivial that M_{i+1} satisfies (I$_{i+1}$) and (II$_{i+1}$) and is maximal with respect to these conditions. Hence we have a unique chain $0 \subset M_1 \subset M_2 \subset \cdots \subset M_i \subset M_{i+1} \subset \cdots$, which is either a finite chain ending with $M_n = L$ or else is an infinite chain. We shall soon rule out the latter possibility.

The natural mapping of L on $K = L/G$ sends the chain of M's into a corresponding chain of normal subloops H_i of K. If we can prove that H_i is properly contained in H_{i+1} for $i = 1, 2, \cdots$, we shall have proved that the H_i cannot form an infinite chain (K being finite) and hence that there is only a finite number of M_i and $M_n = L$. Suppose, then, that $H_k = H_{k+1}$

[4] The concept of central nilpotence is treated by Bruck [3], who gives the name "central series" to series such as $0 \subset H_1 \subset H_2 \subset \cdots \subset H_n = K$. Our condition, "$H_i/H_{i-1}$ in the center of K/H_{i-1}" is clearly equivalent to Bruck's equation (4.2) if the latter is read $H_\pi(K_{i+1}, G) \leqq K_i$. Presumably equation (4.2) as it stands is a misprint.

Of course, we have also exhibited a central series for L so that L is centrally nilpotent and K is *a fortiori* centrally nilpotent [3, Theorem 4C, p. 267].

for some $k \geqq 1$. The complete inverse image of H_k in L is $M_k + G$ and we have $M_k + G = M_{k+1} + G$. Mapping these sets into L/M_{k+1}, we notice that the image of $M_k + G$ is contained in the center of L/M_{k-1}; for the image of an element of M_k is in the center by (II_k), the image of an element of G is in the center since G is in the center of L, and the center of L/M_{k-1} is a loop. Then the image of $M_{k+1} + G$ is in the center of L/M_{k-1} and in particular the image of M_{k+1}, namely M_{k+1}/M_{k-1}, is in the center of of L/M_{k-1}. Moreover, M_{k+1} satisfies (I_k). Hence, by the maximality of M_k, $M_{k+1} \subset M_k$, $M_{k+1} = M_k$, a contradiction. This completes the proof.

In [6] it is proved that if G is an archimedean-ordered group, then so also is L. We can obtain this result as a special case of Theorem 4, once we have the following plausible theorem.

THEOREM 5. *There is a one-to-one correspondence between the l-ideals of G and the l-ideals of L, under which inclusion is preserved; in the terminology of Hahn, the loops L and G have the same order-type.*

Proof. If M is an l-ideal of L, define the corresponding l-ideal $J(M)$ of G as the intersection $M \cap G$. If J is an l-ideal of G, define the corresponding $M(J)$ as the set of all x in L such that, for some y and z in J, $y \leqq x \leqq z$. By Lemma 7, $M(J)$ is the l-ideal generated by J. If M_0 is an l-ideal of L, then $M[J(M_0)] \subset M_0$, clearly; on the other hand, if x is an element of M_0 then for some positive integer N, $xR_x{}^N$ is in G (due to the finiteness of L/G), and so x, being between 0 and $xR_x{}^N$, is in $M[J(M_0)]$. Next, if J_0 is an l-ideal of G, $J[M(J_0)] = M(J_0) \cap G \supset J_0$; and if x is an element of $M(J_0) \cap G$ then x is in G and lies between two elements of J_0 and so is in the l-ideal J_0. This proves the one-to-one correspondence. The rest of Theorem 5 is then obvious.

Now suppose that G is archimedean-ordered, that is, G has no proper l-ideals. Then by Theorem 5, L has no proper l-ideals. In particular, in the chain $L = M_n \supset \cdots \supset M_1 = 0$ of Theorem 4, necessarily $n = 1$ and $L = M_1$ is abelian, has no l-ideals, and so is archimedean-ordered.

4. Ordered loops as loop extensions. A loop L is said to be a loop extension of the loop G by the loop K in case L has a normal subloop G' isomorphic with G and L/G' is isomorphic with K. If G' is in the center of L, then L is a central extension of G by K. We shall often identify G and G', K and L/G'

If L is a loop extension of G by K we can, as usual, put L into one-to-one correspondence with the set L' of all pairs (k, g) with k in K and g in G.

This is done by choosing a "set of representatives" $r(k)$: a single-valued function on K to L with the property that $r(k)$ maps into k modulo G. (We shall always normalize the function $r(k)$ so that $r(0) = 0$.) If x is in L and its residue class modulo G is k_x, then $g_x = x - r(k_x)$ is in G, $x = g_x + r(k_x)$, and the k_x and g_x thus defined are unique. It is easy to see that the mapping

$$x \rightarrow (k_x, g_x) \tag{1}$$

of L into L' is one-to-one and exhausts L'. However, in order that (1) be an isomorphism of L and L' we must define addition in L' in a rather uncomfortable way:

$$(k, g) + (k', g') = (k + k',\ g + g' + f)$$

where f is a suitable function of k, k', g, and g' with certain invertibility properties. But when the extension is central, the function f depends only on k and k' as in the associative case (except that now f need not satisfy any associativity conditions). The function $f(k, k')$ is called a factor set.

LEMMA 11. *If L is a central loop extension of a loop G by a loop K, the correspondence* (1) *between L and L' set up by the set of representatives $r(k)$ is an isomorphism when addition in L' is defined by*

$$(k, g) + (k', g') = (k + k', g + g' + f[k, k']) \tag{2}$$

with a suitable factor set $f(k, k')$. Moreover, since $r(0) = 0$,

$$f(k, 0) = f(0, k') = 0 \tag{3}$$

and G corresponds to the set of all $(0, g)$.

Conversely, if G is an abelian group, if K is a loop and L' is the set of pairs (k, g) with addition defined by (2), *using any factor set f satisfying* (3), *then L' is a central loop extension of G by K.*

The proof of Lemma 11 is the standard one.

If $n = 1$ in Theorem 4, we have the situation described in the following theorem.

LEMMA 12. *Let L be an ordered abelian group, let G be a subgroup of L and $K = L/G$ be finite of order m, so that K is a direct sum of cyclic groups $(k_1) + (k_2) + \cdots + (k_t)$ with the order of k_i equal to m_i. Then there are t elements $g_1, \cdots, g_t$ in G such that*

$$mg + \sum_{i=1}^{t} (m/m_i) j_i g_i = 0 \text{ implies } g = 0,\ j_1 = \cdots = j_t = 0 \tag{4}$$

whenever g is in G and $0 \leqq j_i < m_i$ for all i. And L is isomorphic with the set L' of all pairs (k, g) with k in K and g in G, addition of pairs being defined as in Lemma 11, with factor set

$$f(k, k') = \sum_{i=1}^{t} e_i g_i.$$

Here e_i is defined as follows: If $k = \Sigma j_i k_i$, $k' = \Sigma j'_i k_i$ with $0 \leqq j_i < m_i$, $0 \leqq j' < m_i$, then e_i is the greatest integer in $(j_i + j'_i)/m_i$ (so that e_i is either 0 or 1).

This isomorphism is an order-isomorphism if we define $(k, g) > 0$ in L' to mean $mg + \Sigma(m/m_i) j_i g_i > 0$ in G.

Conversely, given an abelian group K and an ordered abelian group G containing elements g_i that satisfy (4), the set L' under the addition and ordering defined above is an ordered abelian group extension of G by K.

Proof. By Lemma 11, L is isomorphic with the set L' of all pairs, provided a suitable factor set is chosen. We may normalize this factor set by arranging that the representatives $r(k)$ satisfy $r(\Sigma j_i k_i) = \Sigma j_i r(k_i)$ whenever $0 \leqq j_i < m_i$ $(i = 1, \cdots, t)$. If $k = \Sigma j_i k_i$ and $k' = \Sigma j'_i k_i$, then $f(k, k') = r(k) + r(k') - r(k + k') = \Sigma(j_i + j'_i) r(k_i) - \Sigma(j_i + j'_i - e_i m_i) r(k_i) = \Sigma e_i m_i r(k_i)$. But since $m_i k_i = 0$, $m_i r(k_i) = g_i$ is in G and $f(k, k') = \Sigma e_i g_i$. That (4) is true is clear since $mg + \Sigma(m/m_i) j_i g_i = m[g + \Sigma j_i r(k_i)] = m[g + r(k)]$ which is zero if and only if $g + r(k) = 0$, which in turn implies $g = 0$, $r(k) = 0$, $k = 0$, $j_i = 0$ for $i = 1, \cdots, t$. Similarly the ordering of L' suggested in the lemma gives rise to an order-isomorphism because we have defined $(k, g) > 0$ in case $m(k, g) > 0$ in G (strictly, in G', the image of G in L').

As for the converse theorem, by Lemma 11, L' is a loop and is clearly commutative since K and G are and $f(k, k') = f(k', k)$. The associative law can be verified by direct computation. The suggested ordering of L' is clearly linear by (4). Since L' is already abelian and since $(k, g) > 0$ if and only if $m(k, g) > 0$ in G, the sum of two positive elements is positive, and L' is an ordered group.

Suppose that K is a centrally nilpotent loop with central series

$$K = H_n \supset H_{n-1} \supset \cdots \supset H_1 \supset H_0 = 0; \tag{5}$$

that is, each H_i is a normal subloop of K and H_i/H_{i-1} is in the center of K/H_{i-1}. Define $H^i = H_i/H_{i-1}$ $(i = 1, \cdots, n)$. These H^i are called the

factor groups of the series (5). Then $H_1 = H^1$, H_2 is an extension of H_1 by the abelian group H^2, etc., so that K is the result of a succession of loop extensions by abelian groups. This statement and its converse may be formalized as follows. Choose sets of representatives $r_i(a_i)$: functions on H^i to H_i such that, modulo H_{i-1}, $r_i(a_i)$ is a_i (as usual, $r_i(0) = 0$). Then every x in K may be written in one and only one fashion as

$$x = r_n(a_n) + \{r_{n-1}(a_{n-1}) + [\cdot\cdot\cdot + r_1(a_1)]\}, \tag{6}$$

so that the a_i are uniquely defined by x. Thus

$$x \rightarrow (a_n, \cdot\cdot\cdot, a_1) \tag{7}$$

is a one-to-one correspondence between K and the set K' of all sequences $(a_n, \cdot\cdot\cdot, a_1)$ with a_i in H^i.

LEMMA 13. *In order that* (7) *be an isomorphism between loops K and K', it is necessary and sufficient that addition in K' be defined as follows. If $x' = (a_n, \cdot\cdot\cdot, a_1)$, $y' = (b_n, \cdot\cdot\cdot, b_1)$ and $x' + y' = (c_n, \cdot\cdot\cdot, c_1)$, then for $i = 1, 2, \cdot\cdot\cdot, n$,*

$$c_i = a_i + b_i + f_i(a_n, \cdot\cdot\cdot, a_{i+1}; b_n, \cdot\cdot\cdot, b_{i+1}), \tag{8}$$

where each f_i is a suitably chosen function of the arguments indicated, taking values in H^i, $f_n = 0$ and

$$f_i(0, \cdot\cdot\cdot, 0; b_n, \cdot\cdot\cdot, b_{i+1}) = f_i(a_n, \cdot\cdot\cdot, a_{i+1}; 0, \cdot\cdot\cdot, 0) = 0. \tag{9}$$

Conversely, given any set of abelian groups H^i and any functions f_i, with range and domain as above and satisfying (9), *the set K' is a centrally nilpotent loop under the addition defined in* (8).

Thus this loop K' of sequences is the most general centrally nilpotent loop.

Proof. Note that $r_1(a_1) = a_1$ is in H_1 and hence in the center of K. Hence we may write $x = m + a_1$, $m = r_n(a_n) + [\cdot\cdot\cdot + r_2(a_2)]$. Similarly, let $y = p + b_1$, $p = r_n(b_n) + [\cdot\cdot\cdot + r_2(b_2)]$ and write $x + y$ in the normal form (6).

$$\begin{aligned} x + y &= (m + p) + (a_1 + b_1) = [\sum_{i=2}^{n} r_i(c_i) + f_1] + (a_1 + b_1), \\ x + y &= \sum_{i=2}^{n} r_i(c_i) + (a_1 + b_1 + f_1), \end{aligned} \tag{10}$$

where f_1 depends only on $a_n, \cdot\cdot\cdot, a_2$ and $b_n, \cdot\cdot\cdot, b_2$. Moreover, if $a_n, \cdot\cdot\cdot, a_2$ are all zero, then $m = 0$ and clearly $f_1 = 0$. Similarly, if $b_n, \cdot\cdot\cdot, b_2$ are

all zero, $f_1 = 0$. Then we have verified the fact that $c_1 = a_1 + b_1 + f_1$ with f_1 as in the lemma. Now consider the centrally nilpotent loop K/H_1 with the shorter central series $K/H_1 = H_n/H_1 \supset H_{n-1}/H_1 \supset \cdots \supset H_2/H_1 \supset H_1/H_1 = 0$. The factor groups $(H_i/H_1)/(H_{i-1}/H_1)$ are isomorphic with the old factor groups H^i in a natural fashion, so the choice of representatives $r_i(a_i)$ already made in K gives a naturally related set of representatives $\bar{r}_i(a_i)$ in K/H_1. If $\bar{x}$ and $\bar{y}$ are the maps in K/H_1 of x and y, then, in the notation of (10), it is clear that $\bar{x} + \bar{y} = \sum_{i=2}^{n} \bar{r}_i(c_i)$. From this fact we deduce by induction the truth of the direct part of Lemma 13. The converse consists merely in straightforward verification of the definitions.

Now we may consider the general case of an ordered loop L containing a group G in its center wtih a finite quotient loop $K = L/G$. Then there are uniquely defined central series for L, K and G:

$$L = M_n \supset M_{n-1} \supset \cdots \supset M_1 \supset M_0 = 0, \tag{11}$$

$$K = H_n \supset H_{n-1} \supset \cdots \supset H_1 \supset H_0 = 0, \tag{12}$$

$$G = F_n \supset F_{n-1} \supset \cdots \supset F_1 \supset F_0 = 0, \tag{13}$$

where the series for L and K are defined in Theorem 4 and we define $F_i = M_i \cap G$ (note that each F_i is an l-ideal of G). We have already denoted the factor groups H_i/H_{i-1} of (12) by H^i. Likewise, let us denote the factor groups of (11) by $M^i = M_i/M_{i-1}$ and of (13) by $F^i = F_i/F_{i-1}$. Note that each M^i is an ordered abelian group which is an extension of F^i by H^i so that Lemma 12 describes the typical method of building M^i from F^i and H^i. Next, Lemma 13 shows how to build L from the M^i (as well as K from the H^i). Conversely, this general construction will always yield a loop L that can be ordered, but it will not necessarily give a loop with a subgroup of finite index. If we put these two types of extensions together a little more carefully, we arrive at the most general ordered loop L of the kind we are considering. We shall prefer at first to consider the extensions in what might be considered an order opposite to that just described.

By Lemma 13, the central series (12) for K together with any sets of representatives $r_i(a_i)$ determine a standard form (6) for elements of K. Choose sets of representatives $r'_i(a_i)$ in L such that $r'_i(a_i)$ is in M_i for a_i in H^i; $r'_i(a_i)$ modulo G is $r_i(a_i)$; and $r'_i(0) = 0$. Then every element x in L is expressible as

$$x = r'_n(a) + \{r'_{n-1}(a_{n-1}) + [\cdots + r'_1(a_1)]\} + g \tag{14}$$

where both g in G and the a_i in H^i are uniquely determined by x. If we think of G as centrally nilpotent with central series (13), we may also write $g = \sum_{i=1}^{n} s_i(u_i)$ where the u_i are in F^i and $s_i(u_i)$ are sets of representatives in G. The correspondence

$$x \rightarrow (a_n, \cdots, a_1; u_n, \cdots, u_1) \tag{15}$$

then establishes a one-to-one correspondence between L and the set L' of all sequences $(a_n, \cdots, a_1; u_n, \cdots, u_1)$ with a_i in H^i and u_i in F^i. We shall attempt to discover what definitions of addition and ordering in L' make (15) an order-isomorphism.

Suppose $x = \Sigma r'_i(a_i) + g$ as in (14) and $y = \Sigma r'_i(b_i) + h$. Then $x + y = \Sigma r'_i(a_i) + \Sigma r'_i(b_i) + (g + h)$, so it remains to write the correspondent, under (15), of $\Sigma r'_i(a_i) + \Sigma r'_i(b_i)$. For suppose we write $x_0 = \Sigma r'_i(a_i)$, $y_0 = \Sigma r'_i(b_i)$ and $x_0 + y_0 \rightarrow (c_n, \cdots, c_1; w_n, \cdots, w_1)$. Then $x_0 + y_0 = \Sigma r'_i(c_i) + \Sigma s_i(w_i)$ and $x + y = \Sigma r'_i(c_i) + [\Sigma s_i(w_i) + g + h]$, which is in the form (14).

Let $x_0 = \Sigma r'_i(a_i) = m + r'_1(a_1)$, $y_0 = \Sigma r'_i(b_i) = p + r'_1(b_1)$ and note that, since $r'_1(a_1)$ and $r'_1(b_1)$ are in M_1 they are in the center of L. Thus, $x_0 + y_0 = (m + p) + r'_1(a_1) + r'_1(b_1)$. Write $m + p$ in the standard form (14):

$$m + p = \sum_{i=2}^{n} r'_i(c_i) + r'_1(f_1) + \sum_{i=2}^{n} s_i(u_i) + s_1(m_1) \tag{16}$$

for some f_1 in H_1 and m_1 in F_1, both f_1 and m_1 depending only on $a_2, \cdots, a_n$ and $b_2, \cdots, b_n$. Writing $r'_1(a_1) + r'_1(b_1)$ in the standard form, we get

$$r'_1(a_1) + r'_1(b_1) = r'_1(a_1 + b_1) + k_1, \tag{17}$$

where k_1 is an element of G depending on a_1 and b_1. We can say even more: k_1 is in M_1 and hence in F_1. In fact, since M_1 is an ordered abelian group that is an extension of F_1 by the finite group H_1, the function $k_1(a_1, b_1)$ is exactly a factor set of the type described in Lemma 12 when the set of representatives r'_1 is suitably chosen. Combining (16) and (17), we have

$$\begin{aligned} x_0 + y_0 = {} & [\sum_{i=2}^{n} r'_i(c_i) + \sum_{i=2}^{n} s_i(u_i)] + [r'_1(f_1) + r'_1(a_1 + b_1)] \\ & + s_1(m_1) + k_1(a_1, b_1) \\ = {} & [\sum_{i=2}^{n} r'_i(c_i) + \sum_{i=2}^{n} s_i(u_i)] + r'_1(f_1 + a_1 + b_1) \\ & + k_1(f_1, a_1 + b_1) + s_1(m_1) + k_1(a_1, b_1). \end{aligned}$$

Thus $x_0 + y_0 \rightarrow (c_n, \cdots, c_2, c_1; w_n, \cdots, w_2, w_1)$ where

$$c_1 = a_1 + b_1 + f_1(a_n, \cdots, a_2; b_n, \cdots, b_2), \tag{18}$$

$$w_1 = k_1(a_1, b_1) + k_1(f_1, a_1 + b_1) + m_1(a_n, \cdots, a_2; b_n, \cdots, b_2). \tag{19}$$

Formulas almost identical with (18) and (19) are valid for $c_2, \cdots, c_n$ and $w_2, \cdots, w_n$. To see this, consider L/M_1, an ordered loop of the same general character as L. In fact, the series corresponding to (11), (12) and (13) are

$$\begin{aligned} L_1 &= L/M_1 = M_n/M_1 \supset \cdots \supset M_2/M_1 \supset M_1/M_1 = 0, \\ K_1 &= K/H_1 = H_n/H_1 \supset \cdots \supset H_2/H_1 \supset H_1/H_1 = 0, \\ G_1 &= G/F_1 = F_n/F_1 \supset \cdots \supset F_2/F_1 \supset F_1/F_1 = 0, \end{aligned}$$

and the factor groups $(M_i/M_1)/(M_{i-1}/M_1)$, $(H_i/H_1)/(H_{i-1}/H_1)$ and $(F_i/F_1)/(F_{i-1}/F_1)$ are respectively isomorphic in a natural way to M^i, H^i and F^i. Then by an argument very similar to that used in Lemma 13, we see that for (15) to be an isomorphism of loops, addition of sequences must be defined thus:

$$\begin{aligned} (a_n, \cdots, a_1; u_n, \cdots, u_1) &+ (b_n, \cdots, b_1; v_n, \cdots, v_1) \\ = (c_n, \cdots, c_1; 0, \cdots, 0) &+ [(0, \cdots, 0; w_n, \cdots, w_1) \\ &+ (0, \cdots, 0; u_n, \cdots, u_1) + (0, \cdots, 0; v_n, \cdots, v_1)] \end{aligned} \tag{20}$$

where

$$c_i = a_i + b_i + f_i(a_n, \cdots, a_{i+1}; b_n, \cdots, b_{i+1}), \tag{21}$$

$$w_i = k_i(a_i, b_i) + k_i(f_i, a_i + b_i) + m_i(a_n, \cdots, a_{i+1}; b_n, \cdots, b_{i+1}) \tag{22}$$

for $i = 1, 2, \cdots, n$, and the addition in brackets in (20) is performed via the isomorphism described in Lemma 13 between G and the set of all $(0, \cdots, 0; u_n, \cdots, u_1)$. Here f_i are the factor sets of Lemma 13 defined by K and the representatives $r_i(a_i)$; k_i are the factor sets as in Lemma 12 determined by M^i when considered as a group extension of F^i by H^i, provided the representatives r_i are suitably chosen; and m_i is another, new factor set. Also, as in Lemma 13, $f_n = 0$, $m_n = 0$, and

$$\begin{aligned} f_i(0, \cdots, 0; b_n, \cdots, b_{i+1}) &= f_i(a_n, \cdots, a_{i+1}; 0, \cdots, 0) = 0, \\ m_i(0, \cdots, 0; b_n, \cdots, b_{i+1}) &= m_i(a_n, \cdots, a_{i+1}; 0, \cdots, 0) = 0. \end{aligned} \tag{23}$$

As for the ordering necessary in the loop L' of sequences, let us write each element x' of L' in the form $(t_n, \cdots, t_1)$, where each t_i represents a pair (a_i, u_i). Such a pair may be considered, in a natural fashion, an element

of the ordered group M^i (Lemma 12). Suppose that $x' = (t_n, \cdots, t_1)$ and that j is the largest index such that $t_j \neq 0$. Then x' is in M_j but not in M_{j-1}; x' modulo M_{j-1} is the element t_j of M^j, and since M_{j-1} is an l-ideal of M_j,

$$(24) \qquad x' > 0 \text{ if and only if } t_j > 0 \text{ in } M^j.$$

Thus the ordering of L' must be lexicographic ordering of the sequences $(t_n, \cdots, t_1)$.

THEOREM 6. *If L is an ordered loop with a subloop G of finite index contained in the center of L, then L is order-isomorphic with the loop L' of all sequences $(a_n, \cdots, a_1;\ u_n, \cdots, u_1)$ with a_i in H^i and u_i in F^i, provided addition in L' is defined by* (20), (21) *and* (22) *and the ordering of L' is the lexicographic ordering* (24).

Conversely, given an ordered abelian group G and a set of finite abelian groups H^i $(i = 1, \cdots, n)$, we can build an ordered loop L' as follows. Find a central series $G = F_n \supset \cdots \supset F_1 \supset 0$ such that each F_i is an l-ideal of G and such that for each i there exists an ordered abelian group M^i which is an extension of $F^i = F_i/F_{i-1}$ by H^i as in Lemma 12. *(If this is not possible, there is no ordered loop extension L of G, of the type we are considering with the H^i as the factor groups of a central series of L/G.) Let the factor set used to define M^i as an extension of F^i by H^i be $k_i(a_i, b_i)$—a function on H^iH^i to F^i—and let L' be the set of all sequences $(a_n, \cdots, a_1; u_n, \cdots, u_1)$ with a_i in H^i and u_i in F^i. If addition in L' is defined by* (20), (21) *and* (22), *the functions f_i and m_i being chosen subject only to* (23), *and if L' is ordered lexicographically as in* (24), *then L' is an ordered loop which is a central loop extension of G by a centrally nilpotent finite loop whose factor groups are $H^n, \cdots, H^1$.*

We have already proved the direct part of this theorem. The converse is merely a matter of verifying definitions.

Theorem 6 essentially says that an ordered central extension of G by K always exists if G has sufficiently many l-ideals and if the corresponding factor groups F^i are not too nearly complete [cf. condition (4), Lemma 12, which asserts the existence in F^i of a number of elements not divisible by the integer m.]. Thus, in a sense, the condition that K be centrally nilpotent is not only necessary (Theorem 4) but also sufficient.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] A. A. Albert, "Quasigroups. I," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 507-519.

[2] G. Birkhoff, "Lattice-ordered groups," *Annals of Mathematics*, vol. 43 (1942), pp. 298-331.

[3] R. H. Bruck, "Contributions to the theory of loops," *Transactions of the American Mathematical Society*, vol. 60 (1946), pp. 245-354.

[4] W. Krull, "Allgemeine Bewertungstheorie," *Journal für die reine und angewandte Mathematik*, vol. 167 (1932), pp. 160-196.

[5] S. Lefschetz, *Algebraic Topology*, American Mathematical Society Colloquium Publications, vol. XXVII, New York, 1942.

[6] D. Zelinsky, "Nonassociative valuations," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 175-183.

APPROXIMATIONS TO A CLASS OF DOUBLE INTEGRALS OF FUNCTIONS OF LARGE NUMBERS.*

By Leetch C. Hsu.

1. Introduction. It is known that the following theorem due to Laplace [1] plays a fundamental rôle in the analytic theory of probabilities:

Let $\phi(x)$, $h(x)$ and $f(x) = e^{h(x)}$ be continuous functions defined on the closed interval $a \leqq x \leqq b$ such that

1) $\phi(x)[f(x)]^n$ is absolutely integrable over (a, b), $n = 0, 1, 2, \cdots$,

2) the first and second derivatives of $h(x)$ exist,

3) $h(x)$ has a maximum value at $x = \xi (a < \xi < b)$ in the absolute sense that $h(x) < h(\xi)$ for all x of (a, b) other than ξ and so that $h'(\xi) = 0$, $h''(\xi) < 0$,

4) $h''(x)$ and $\phi(x)$ are continuous at $x = \xi$, where $\phi(\xi) \neq 0$.

Then as n is very large we have the asymptotic formula

$$\int_a^b \phi(x)[f(x)]^n dx \sim \phi(\xi)[f(\xi)]^{n+\frac{1}{2}}\{-2\pi/nf''(\xi)\}^{\frac{1}{2}}.$$

This theorem involves many applications, e. g. asymptotic formulas of Stirling, Wallis and for Legendre's polynomial of the n-th degree are all obtainable from it. Its connection with a class of series expansions had been discussed more completely by Darboux [2]. Other interesting examples may be found in the book of Pólya and Szegö [3].

In this paper we shall prove three theorems concerning a class of double integrals of functions of large numbers. Let $\phi(x, y)$, $h(x, y)$ and $f(x, y) = \exp(h(x, y))$ be continuous functions defined on a closed region S and satisfying certain required conditions. It is the object of this paper to investigate the asymptotic behavior of the integral

$$\iint_S \phi(x, y)[f(x, y)]^n dS.$$

Throughout the paper it will be assumed that $f(x, y)$ or $h(x, y)$ has a

* Received August 27, 1947.

maximum value at a point (x_0, y_0) of S in the sense that $f(x, y) < f(x_0, y_0)$ for all points of S other than (x_0, y_0). If there are two or more than two points at which the function assumes the same maximum value, we may divide S into subregions so that the same consideration is applicable.

Since (x_0, y_0) may be a boundary point or an interior point of S, two cases should be treated separately. Moreover, if it is a boundary point of S, there are also two important cases to be considered. It will be seen later that the asymptotic formulas for all these cases are different.

2. Theorems and proofs. We shall now state and prove three theorems. The first and third theorems are concerned with the case where (x_0, y_0) is a boundary point of S. The statement of Theorem 1 is as follows:

THEOREM 1. *Let $\phi(x,y)$, $h(x,y)$ and $f(x,y) = \exp(h(x,y))$ be functions of (x,y) continuous together with f_x, f_{xx}, f_{xy}, f_y, f_{yy} within and on the boundary C (analytic curve) of a region S such that*

1) *$\phi(x,y)[f(x,y)]^n$ is absolutely integrable over S $(n = 0, 1, 2, \cdots)$,*

2) *$f(x,y)$ has a maximum value at a point of the boundary C (in the absolute sense), (x_0, y_0) say, such that the first and second directional derivatives of f with respect to the arc length s of C are respectively $= 0$ and < 0 at that point,*

3) *$\phi(x,y)$ and $f_x{}^2 + f_y{}^2$ are not zero at (x_0, y_0): $\phi(x_0, y_0) \neq 0$, $(f_x)_0{}^2 + (f_y)_0{}^2 \neq 0$.*
Then the double integral of $\phi(x,y)[f(x,y)]^n$ taken over S is asymptotic to

$$\frac{\{2\pi\}^{\frac{3}{2}}\phi(x_0,y_0)[f(x_0,y_0)]^{n+(3/2)}}{\{n^3[K(f_x{}^2+f_y{}^2)_0{}^{3/2} - ((f_x)_0 D_y - (f_y)_0 D_x)_0{}^{(2)} f]\}^{\frac{1}{2}}},$$

where K is the curvature of C at (x_0, y_0) and $((f_x)_0 D_y - (f_y)_0 D_x)_0{}^{(2)} f$ denotes symbolically $(f_x)_0{}^2 (f_{yy})_0 - 2(f_x)_0 (f_y)_0 (f_{xy})_0 + (f_y)_0{}^2 (f_{xx})_0$.

Proof. Let s be the arc length measured from (x_0, y_0) in the positive sense along C. Then $(df/ds)_0$ is the directional derivative of $f(x,y)$ along the direction of the positive tangent to C at (x_0, y_0). Since $f(x,y)$ has a maximum value at (x_0, y_0) we have

$$(1) \qquad (df/ds)_0 = (\partial f/\partial x)_0 (dx/ds)_0 + (\partial f/\partial y)_0 (dy/ds)_0 = 0.$$

It follows that

$$(2.1) \qquad (dx/ds)_0 = \pm (f_y)_0 / \{(f_x)_0{}^2 + (f_y)_0{}^2\}^{\frac{1}{2}} = \cos\theta,$$

$$(2.2) \qquad (dy/ds)_0 = \mp (f_x)_0 / \{(f_x)_0{}^2 + (f_y)_0{}^2\}^{\frac{1}{2}} = \sin\theta,$$

where θ is the direction angle measured from the positive x-axis to the tangent to C at (x_0, y_0). The sign $+$ or $-$ in (2.1) is taken according as $(\partial f/\partial y)_0 (dx/ds)_0 > 0$ or < 0; and correspondingly the sign $-$ or $+$ in (2.2) as $(\partial f/\partial x)_0 (dy/ds)_0 < 0$ or > 0.

Let the origin of the coordinate system be translated to (x_0, y_0) and then turn the system through an angle θ. We thus have the new coordinates (X, Y) for a point of C so that $x - x_0 = X\cos\theta - Y\sin\theta$ and $y - y_0 = X\sin\theta + Y\cos\theta$. By the well-known formula of Frenet, we may expand X, Y in power series of s about $s = 0$:

$$(3)\qquad \begin{cases} X = (s/1) - (s^3/6R^2) + \cdots, \\ \pm Y = (s^2/2R) - s^3/6R^2(dR/ds)_0 + \cdots. \end{cases}$$

where the sign $+$ or $-$ of Y is taken according as the curve C lies to the left or right of its tangent at $s = 0$, i.e., according as the curvature $K > 0$ or < 0 at that point. Now since the coordinates of a point (x, y) on the curve C are functions of s, we may write $f(x, y) = \Psi(s)$ so that $f(x_0, y_0) = \Psi(0)$ and we have

$$(4)\quad \begin{aligned} d\Psi/dS = \partial f/\partial x[(dX/dS)\cos\theta - (dY/dS)\sin\theta] \\ + \partial f/\partial y[(dX/dS)\sin\theta + (dY/dS)\cos\theta]. \end{aligned}$$

Differentiating this expression and putting $s = 0$ we obtain, in view of (3) and (2),

$$(5)\quad \begin{aligned} (d^2\Psi/dS^2)_0 &= [(\partial/\partial x)_0\cos\theta + (\partial/\partial y)_0\sin\theta]^{(2)}f \\ &\quad + K[(\partial f/\partial y)_0\cos\theta - (\partial f/\partial x)_0\sin\theta] \\ &= \pm K(f_x^2 + f_y^2)_0^{1/2} + (f_x^2 + f_y^2)_0^{-1}[(f_x)_0(\partial/\partial y)_0 \\ &\quad - (f_y)_0(\partial/\partial x)_0]^{(2)}f, \end{aligned}$$

where $[(f_x)_0(\partial/\partial y)_0 - (f_y)_0(\partial/\partial x)_0]^{(2)}f$ denotes symbolically $(f_x)_0^2(f_{yy})_0 - 2(f_x)_0(f_y)_0(f_{xy})_0 + (f_y)_0^2(f_{xx})_0$ and $K = \pm 1/R$ is the curvature of C at (x_0, y_0). The prefixed sign $+$ or $-$ of K is chosen in accordance with that of (2.1). But since we have taken a positive sense for the measurement of C and since the function $f(x, y)$ has a maximum value at (x_0, y_0), it is easy to observe that for each case

$$(\partial f/\partial y)_0(dx/ds)_0 < 0 \text{ or correspondingly } (\partial f/\partial x)_0(dy/ds)_0 > 0.$$

This fact can also be justified by comparing the two sides of (7) (see below). Thus it is seen that the sign of K must be negative, i.e. we may definitely replace $\pm K$ by $-K$ in the expression (5).

Let $\Phi(s)=\phi(x,y)$, $\Phi(0)=\phi(x_0,y_0)$, where (x,y) is a moving point of C. Then by Laplace's theorem we obtain

$$(6)\qquad \int_C \phi(x,y)[f(x,y)]^n ds = \int_{s_0}^{s_1} \Phi(s)[\Psi(s)]^n ds$$
$$= \Phi(0)[\Psi(0)]^{n+\frac{1}{2}}\{(-2\pi/n\Psi''(0))(1+\epsilon)\}^{\frac{1}{2}},$$

where s_1-s_0 is the total arc length of C and $\epsilon\to 0$ as $n\to\infty$.

By hypothesis 3), the values of $f_x(x_0,y_0)$ and $f_y(x_0,y_0)$ are not both zero, $f_y(x_0,y_0)\neq 0$ say, so that by Green's theorem and (6) we have

$$-(n+1)\iint_S [f(x,y)]^n f_y(x,y)\,dS = \int_C [f(x,y)]^{n+1}dx$$
$$(7)\qquad = (dx/ds)_0[\Psi(0)]^{n+3/2}(-2\pi/n\Psi''(0))^{\frac{1}{2}}(1+\epsilon_n)$$
$$=\pm [f(x_0,y_0)]^{n+3/2}\{-2\pi(f_y)_0{}^2/n[-K(f_x{}^2+f_y{}^2)_0{}^{3/2}$$
$$+((f_x)_0D_y-(f_y)_0D_x)_0{}^{(2)}f]\}^{\frac{1}{2}}(1+\epsilon_n),$$

where the sign $+$ or $-$ in the last expression is taken according as $f_y(x_0,y_0)<0$ or >0 and $\epsilon_n\to 0$ as $n\to\infty$.

Let ϵ be an arbitrarily small number >0. There can be found a neighborhood of (x_0,y_0) in S, σ say, such that

$$|f_y(x,y)-f_y(x_0,y_0)|<\epsilon \qquad \text{for } (x,y)\,\varepsilon\,\sigma.$$

Then by the mean value theorem in the integral calculus, we may write

$$I=\iint_S \{f(x,y)/f(x_0,y_0)\}^n f_y(x,y)\,dS$$
$$= f_y(x',y')\iint_\sigma \{f(x,y)/f(x_0,y_0)\}^n dS + 0(k^n),$$

where (x',y') is an interior point of σ and $0<k<1$, in fact,

$$k=\mathrm{Max}\,\{f(x,y)/f(x_0,y_0)\}<1 \qquad \text{for } (x,y)\,\not\varepsilon\,\sigma.$$

Hence it is seen that the integral I is equal to

$$(f_y(x_0,y_0)+\delta)\iint_\sigma \{f(x,y)/f(x_0,y_0)\}^n dS + 0(k^n); \qquad |\delta|<\epsilon.$$

Since ϵ is arbitrary we may conclude that

$$\iint_S [f(x,y)]^n f_y(x,y)\,dS \sim f_y(x_0,y_0)\iint_S [f(x,y)]^n dS.$$

A comparison with (7) gives at once the result

$$\iint_S [f(x,y)]^n dS \sim [f(x_0,y_0)]^{n+3/2}\{-2\pi/(n^3[-K(f_x^2+f_y^2)_0^{3/2} + ((f_x)_0(D_y)_0-(f_y)_0(D_x)_0)^{(2)}f])\}^{\frac{1}{2}}.$$

Using hypothesis 3) and the mean value theorem again, it can be shown in a similar manner that

$$\iint_S \phi(x,y)[f(x,y)]^n dS \sim \phi(x_0,y_0)\iint_S [f(x,y)]^n dS.$$

Hence the theorem is proved.

THEOREM 2. *Let $\phi(x,y)$, $h(x,y)$ and $f(x,y)=\exp(h(x,y))$ be continuous functions defined on S such that*

1) $\phi(x,y)[f(x,y)]^n$ *is absolutely integrable over S $(n=0,1,2,\cdots)$,*

2) $f_x, f_{xx}, f_y, f_{yy}, f_{xy}$ *exist and are continuous throughout S,*

3) $h(x,y)$ *has an absolute maximum value at an interor point (x_0,y_0) of S so that $(h_x)_0=(h_y)_0=0$, $(h_{xx}h_{yy})_0-(h_{xy})_0^2>0$,*

4) $\phi(x,y)$ *is continuous at (x_0,y_0) and $\phi(x_0,y_0)\neq 0$.*

Then the double integral of $\phi(x,y)[f(x,y)]^n$ taken over S is asymptotic to

$$2\pi\phi(x_0,y_0)[f(x_0,y_0)]^n/n\{h_{xx}(x_0,y_0)h_{yy}(x_0,y_0)-h_{xy}^2(x_0,y_0)\}^{\frac{1}{2}}.$$

Proof. We now consider the integral

$$I_1=\iint_S \{f(x,y)/f(x_0,y_0)\}^n dS=\iint_S e^{n[h(x,y)-h(x_0,y_0)]}dS.$$

By hypothesis 3) and Taylor's expansion we have

$$h(x,y)-h(x_0,y_0)=\tfrac{1}{2}h_{xx}(\xi,\eta)(x-x_0)^2 + h_{xy}(\xi,\eta)(x-x_0)(y-y_0)+\tfrac{1}{2}h_{yy}(\xi,\eta)(y-y_0)^2,$$

where ξ and η are interior points belonging to the intervals (x_0,x) and (y_0,y) respectively, so they are also functions of x and y. It is clear that $h_{xx}(x_0,y_0)+2h_{xy}(x_0,y_0)t+h_{yy}(x_0,y_0)t^2$ is a negative definite quadratic function of t and that $h_{xx}(\xi,\eta)\to h_{xx}(x_0,y_0)$, $h_{xy}(\xi,\eta)\to h_{xy}(x_0,y_0)$, $h_{yy}(\xi,\eta)\to h_{yy}(x_0,y_0)$ as $(x,y)\to(x_0,y_0)$. Thus on writing $x-x_0=X$, $y-y_0=Y$ and making $(X,Y)\to(0,0)$ along any direction, $Y=t\cdot X$ say, we have

$$\lim_{(X,Y)\to(0,0)} \frac{h_{xx}(\xi,\eta)X^2+2h_{xy}(\xi,\eta)XY+h_{yy}(\xi,\eta)Y^2}{h_{xx}(x_0,y_0)X^2+2h_{xy}(x_0,y_0)XY+h_{yy}(x_0,y_0)Y^2}=1.$$

Hence we may write

$$\begin{aligned} h_{xx}(\xi,\eta)X^2+2h_{xy}(\xi,\eta)XY+h_{yy}(\xi,\eta)Y^2 \\ = (\alpha X+2\beta XY+\gamma Y^2)(1+\zeta(X,Y)), \end{aligned}$$

where $\alpha=h_{xx}(x_0,y_0)$, $\beta=h_{xy}(x_0,y_0)$, $\gamma=h_{yy}(x_0,y_0)$ and $\zeta(X,Y)\to 0$ as $(X,Y)\to(0,0)$.

Let θ be an angle such that when the (X,Y)-coordinate system is turned through it the quadratic form $\alpha X^2+2\beta XY+\gamma Y^2$ will be reduced to $\lambda_1 u^2+\lambda_2 v^2$. By the well-known invariants for conic sections it is seen that λ_1, λ_2 are roots of the equation $\lambda^2-(\alpha+\gamma)\lambda-(\beta^2-\alpha\gamma)=0$, so that $\lambda_1\lambda_2=h_{xx}(x_0,y_0)h_{yy}(x_0,y_0)-h^2_{xy}(x_0,y_0)$, and we may rewrite $(\alpha X^2+2\beta XY+\gamma Y^2)(1+\zeta(X,Y))$ as $(\lambda_1u^2+\lambda_2v^2)(1+\rho(u,v))$, where $\rho(u,v)\to 0$ as $(u,v)\to(0,0)$.

Let ϵ be an arbitrarily small number >0 and <1. There can be found a positive number δ such that $|\rho(u,v)|<\epsilon$ whenever $u^2+v^2\leq\delta^2$. And let R be a circular region with radius δ and with $(0,0)$ as its center. Then since the transformation of the (X,Y)-system to the (u,v)-system is a rotation we have

$$\begin{aligned} I_1 &= \iint_R \exp[(n/2)(\alpha X^2+2\beta XY+\gamma Y^2) \\ &\qquad \times (1+\zeta(X,Y))]dXdY+0(k_1{}^n) \qquad (0<k_1<1) \\ &= \int_{-\delta/\sqrt{2}}^{\delta/\sqrt{2}}\int_{-\delta/\sqrt{2}}^{\delta/\sqrt{2}} \exp[(n/2)(\lambda_1u^2+\lambda_2v^2) \\ &\qquad \times (1+\rho(u,v))]dudv+0(k_2{}^n) \qquad (0<k_2<1) \end{aligned} \tag{8}$$

Therefore it is sufficient to consider the integral on the right-hand side of (8). Denote the integral by I_2 and let $\operatorname{Max}\rho(u,v)=M(\delta)$, $\operatorname{Min}\rho(u,v)=m(\delta)$ for $u^2+v^2\leq\delta^2$. Then clearly

$$I_2<\left(\int_{-\delta/\sqrt{2}}^{\delta/\sqrt{2}} e^{\frac{1}{2}n\lambda_1(1+m(\delta))u^2}du\right)\left(\int_{-\delta/\sqrt{2}}^{\delta/\sqrt{2}} e^{\frac{1}{2}n\lambda_2(1+m(\delta))v^2}dv\right). \tag{9}$$

Noticing that $\lambda_1<0$, $\lambda_2<0$ we see by Laplace's theorem that the right-hand side of (9) is asymptotic to

$$\{-2\pi/(1+m(\delta))n\lambda_1\}^{\frac{1}{2}}\{-2\pi/(1+m(\delta))n\lambda_2\}^{\frac{1}{2}}.$$

In a similar manner we obtain, as n is large enough,

$$I_2 > \{-2\pi/(1+M(\delta))n\lambda_1\}^{\frac{1}{2}}\{-2\pi/(1+M(\delta))n\lambda_2\}^{\frac{1}{2}}.$$

Since ϵ is arbitrary and $|M(\delta)| < \epsilon$, $|m(\delta)| < \epsilon$ we obtain that I_2 is asymptotic to

$$\{-2\pi/n\lambda_1\}^{\frac{1}{2}}\{-2\pi/n\lambda_2\}^{\frac{1}{2}} = 2\pi/n\{h_{xx}(x_0,y_0)h_{yy}(x_0,y_0) - h^2{}_{xy}(x_0,y_0)\}^{\frac{1}{2}}.$$

Now it is clear $I_1 \sim I_2$. Hence it follows that

$$\iint_S [f(x,y)]^n dS \sim 2\pi[f(x_0,y_0)]^n/n\{h_{xx}(x_0,y_0)h_{yy}(x_0,y_0) - h^2{}_{xy}(x_0,y_0)\}^{\frac{1}{2}}.$$

The theorem thus follows by the mean value theorem for integral calculus.

It is easily verified that $h_{xx}(x,y)h_{yy}(x,y) - h_{xy}{}^2(x,y)$ is an invariant form under all unimodular linear transformations in two dimensions. More generally we may obtain an invariant form as follows: Let there be given a 1-to-1 and continuous mapping $x = \Psi_1(X,Y)$, $y = \Psi_2(X,Y)$ so that $\phi(x,y) = \Phi(X,Y)$, $h(x,y) = H(X,Y)$, $f(x,y) = F(X,Y)$ and (x_0,y_0) corresponds to (X_0,Y_0), where the Jacobian of Ψ_1, Ψ_2 with respect to X, Y remains of constant sign in a neighborhood of (x_0,y_0). Then the double integral in question is asymptotic to

$$\frac{2\pi\Phi(X_0,Y_0)[F(X_0,Y_0)]^n}{n\{H_{XX}(X_0,Y_0)H_{YY}(X_0,Y_0) - H^2{}_{XY}(X_0,Y_0)\}^{\frac{1}{2}}}(\partial(x,y)/\partial(X,Y))_{\substack{x=x_0\\ y=y_0}}.$$

Upon comparing we get

$$(h_{xx}h_{yy} - h^2{}_{xy})_0 = (H_{XX}H_{YY} - H^2{}_{XY})_0(\partial(X,Y)/\partial(x,y))_0{}^2$$

Thus an invariant form can be written symbolically and symmetrically as follows:

$$(10) \qquad (h_{xx}h_{yy} - h^2{}_{xy})_0(\partial(x,y))_0{}^2 = (H_{XX}H_{YY} - H^2{}_{XY})_0(\partial(X,Y))_0{}^2.$$

In particular, if the mapping is a unimodular linear transformation, we have $h_{xx}h_{yy} - h^2{}_{xy} \equiv H_{XX}H_{YY} - H^2{}_{XY}$.

The next problem to be considered in this paper may be stated as follows: Let there be given all conditions of Theorem 2 and let C be a curve passing through the point (x_0,y_0) so that the region S is divided into two subregions, S_1 and S_2 say, where C is the common boundary of S_1 and S_2. Then we ask the following question: What is the asymptotic value of

$$\iint_{S_1} \phi(x,y)[f(x,y)]^n dS \quad \text{or} \quad \iint_{S_2} \phi(x,y)[f(x,y)]^n dS?$$

Clearly this question has not been answered by Theorem 1, because in the present case the condition 3) of the theorem is not satisfied. Thus in answering the question we are led to prove the following theorem.

THEOREM 3. *Let $\phi(x,y)$, $h(x,y)$ and $f(x,y) = \exp(h(x,y))$ be continuous functions defined on a region S such that all the conditions of Theorem 2 are fulfilled. Let C be an analytic curve passing through the point (x_0, y_0) such that the region S is divided into two subregions, S_1 and S_2 say. Then the integral $\iint \phi(x,y)[f(x,y)]^n dS$ taken over either of the regions S_1 or S_2 is asymptotic to*

$$\pi\phi(x_0,y_0)[f(x_0,y_0)]^n/n\{h_{xx}(x_0,y_0)h_{yy}(x_0,y_0) - h^2_{xy}(x_0,y_0)\}^{\frac{1}{2}}.$$

Proof. We shall need the following lemma: If L is a straight line passing through the point (x_0, y_0) and if S_1 and S_2 are subregions of S divided by the line, then the integral $\iint \phi(x,y)[f(x,y)]^n dS$ taken over either of S_1 and S_2 is asymptotic to

$$\tfrac{1}{2} \iint_{S} \phi(x,y)[f(x,y)]^n dS.$$

The proof of the lemma is simple. Since $h_{xx}(x,y)h_{yy}(x,y) - h^2_{xy}(x,y)$ is an invariant form under unimodular linear transformations, there is no loss of generality in assuming that L is parallel to the x-axis. Then it is easy to show that the integral of $\phi(x,y)[f(x,y)]^n$ taken over S_1 is asymptotically equal to that over S_2.

To see this it is sufficient to consider the right-hand side of (8). Let R be a circular region with radius δ and with (x_0, y_0) as its center and let T be a square inscribed in R (see Figure 1). Then it is clear from the proof of Theorem 2 that

$$I_1 \sim \iint_R \exp[(n/2)(\alpha X^2 + 2\beta XY + \gamma Y^2)]dXdY$$

$$= \iint_T \exp[(n/2)(\alpha X^2 + 2\beta XY + \gamma Y^2)]dXdY + 0(k^n)$$

$$= \left(\int_{-\delta/\sqrt{2}}^{\delta/\sqrt{2}} e^{\frac{1}{2}n\lambda_1 u^2}du\right)\left(\int_{-\delta/\sqrt{2}}^{\delta/\sqrt{2}} e^{\frac{1}{2}n\lambda_2 v^2}dv\right) + 0(k^n),$$

where $0 < k < 1$. Since the integral $\iint \exp[(n/2)(\alpha X^2 + 2\beta XY + \gamma Y^2)]dXdY$

taken over σ_2 has the same value as that over σ_1, (Fig. 1) we see that the integral taken over S_2 is asymptotic to

$$\left(\int_0^{\delta/\sqrt{2}} e^{\frac{1}{2}n\lambda_2 v^2}dv\right)\left(\int_{\delta/\sqrt{2}}^{-\delta/\sqrt{2}} e^{\frac{1}{2}n\lambda_1 u^2}du\right) \sim \tfrac{1}{2}I_1.$$

Hence the lemma is established in a manner similar to that used in the proof of Theorem 2.

We may now complete the proof of the theorem as follows: Let C be an analytic curve passing through the point $P_0(x_0, y_0)$ and let L be a straight

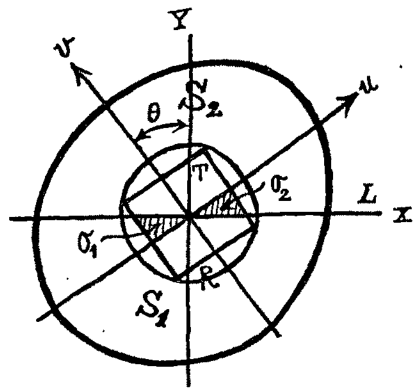


FIG. 1.

line tangent to C at P_0. In order to apply the lemma just proved we have to construct a 1-to-1 continuous mapping A: $X = \phi_1(s, t)$, $Y = \phi_2(s, t)$ such that (i) the S-region is mapped onto a Σ-region of the (s, t)-plane; (ii) the curve C in S is mapped into a straight line l in Σ; (iii) the origin is mapped into the origin. For such a purpose, we may assume that the curve C has been expressed as a function of the arc length s, viz. $X = \psi_1(s)$, $Y = \psi_2(s)$, in particular, for a neighboring point (X, Y) of P_0, we have (cf. (3))

$$\text{(11)} \qquad \begin{cases} X = s/1 - s^3/6R^3 + \cdots = \psi_1(s), \\ Y = \pm\,[s^2/2R - s^3/6R^2(dR/dS)_0 + \cdots] = \psi_2(s), \end{cases}$$

where the arc length s is measured from (x_0, y_0) in a positive sense. Then a required mapping A may be written simply

$$A: \qquad X = \psi_1(s), \qquad Y = \psi_2(s) + t.$$

Clearly this mapping satisfies all conditions required, e.g. the curve C is mapped by it into a straight line $t = 0$ in the (s, t)-plane (see Figure 2).

Moreover, if we write $h(x,y) = H(X,Y) = H^*(s,t)$, then we easily obtain

$$H^*_{ss}(s,t) = H_{XX}(X,Y)\,(\psi'_1(s))^2 + 2H_{XY}(X,Y)\psi'_1(s)\psi'_2(s)$$
$$\qquad + H_{YY}(X,Y)\,(\psi'_2(s))^2 + H_X(X,Y)\psi''_1(s) + H_Y(X,Y)\psi''_2(s),$$
$$H^*_{st}(s,t) = H_{XY}(X,Y)\psi'_1(s) + H_{YY}(X,Y)\psi'_2(s),$$
$$H^*_{tt}(s,t) = H_{YY}(X,Y).$$

so that by (11) we have

$$H^*_{st}(0,0) = H_{XX}(0,0) + (1/R)H_Y(0,0) = H_{XX}(0,0),$$
$$(H_Y(0,0) = 0)$$
$$H^*_{st}(0,0) = H_{XY}(0,0), \qquad H^*_{tt}(0,0) = H_{YY}(0,0).$$

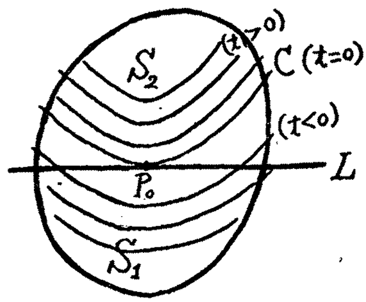


FIG. 2.

It follows that

$$\begin{aligned} & H^*_{ss}(0,0)H^*_{tt}(0,0) - H^{*2}_{st}(0,0) \\ (12) \qquad & \qquad = H_{XX}(0,0)H_{YY}(0,0) - H^2_{XY}(0,0) \\ & \qquad = h_{xx}(x_0,y_0)h_{yy}(x_0,y_0) - h^2_{xy}(x_0,y_0). \end{aligned}$$

The relation (12) can also be justified by (10) and the fact that $(\partial(X,Y)/\partial(s,t))_0 = 1$. Hence by the lemma and Theorem 2 the theorem is proved.

For the case where (x_0, y_0) is an interior point of S, there are also sub-cases not considered. For example, (x_0, y_0, z_0) may be a point of the surface $z = f(x,y)$ at which no horizontal tangent plane exists. In particular, (x_0, y_0, z_0) may be a conic point of the surface. But for such a case the asymptotic behavior of the double integral in question seems simpler. In fact, it may be shown without difficulty that the asymptotic value for this case is much smaller than those we have considered, but we shall not discuss it in this paper.

3. Geometrical interpretations. We shall now in this last section explain the asymptotic formulas obtained in **2** geometrically. It is evident that the asymptotic behavior of the double integral depends essentially on

the neighborhood of the point (x_0, y_0) at which the function $f(x,y)$ takes its maximum. Theorem 1 shows that the asymptotic value of the integral depends on the curvature of the boundary at (x_0, y_0). For definiteness, we may think of the boundary as a path of integration in the positive sense so that the curvature K will be ≥ 0 or ≤ 0 according as the curve lies to the left or right of its tangent at (x_0, y_0). Now we may draw a small circle with (x_0, y_0) as its center. In this way, it is seen that the intersecting area of the circle and the region S will be $\leq$ the area of the semicircle cut out by the tangent if $K \geq 0$; and will be $\geq$ the area of the semicircle if $K \leq 0$. Thus clearly the asymptotic value for the first case should be $\leq$ the value for the second case. And this agrees with the fact that the factor K occurs in the denominator:

$$\{n^3[K(f_x^2 + f_y^2)_0{}^{3/2} - ((f_x)_0 D_y - (f_y)_0 D_x)_0{}^{(2)} f]\}^{\frac{1}{3}}.$$

In particular, if the boundary C of S is a convex closed curve (i. e. $K \geq 0$ for all points of C), then the asymptotic value will be a maximum when and only when $K = 0$ at the point (x_0, y_0).

Theorem 2 shows that the asymptotic value in its case is much greater than that given by Theorem 1, in fact, the orders of their denominators are respectively $0(n^{-1})$ and $0(n^{-3/2})$. The reason may be that in Theorem 2 the neighboring points of (x_0, y_0, z_0) on the surface $z = f(x,y)$ have infinitesimal distances of order 2 from the tangent plane $z = z_0$, but the situation is not so for the case of Theorem 1.

For the reasoning of Theorem 3, we may draw a small circle (of radius δ) with (x_0, y_0) as its center so that it is divided into two portions by the curve C. Since C has a tangent at (x_0, y_0), we see that the areas of these two portions are asymptotically equivalent as $\delta \to 0$. Thus it is geometrically evident that the two double integrals of $\phi(x,y)[f(x,y)]^n$ taken over S_1 and S_2 are also asymptotically equal.

TSING-HUA UNIVERSITY, PEIPING, CHINA.

BIBLIOGRAPHY.

[1] P. S. Laplace, *Théorie analytique des probabilités*, Part 2, Chapter 1, p. 89, Oeuvres, Paris: Gauthier-Villars (1886).

[2] G. Darboux, "Mémoire sur l'approximation des fonctions de très-grands nombres, et sur une classe étendue de développements en série," *Journal de Mathématiques*, Série 3, Vol. 4, pp. 5-56, 377-416 (1878).

[3] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, New York, Chapter 5, pp. 77-82 (1945).

ON BESSEL SUMMATION.*

By K. CHANDRASEKHARAN and OTTO SZÁSZ.

We present in this paper[1] some results concerning Cesàro and Bessel summability; each method contains a parameter, representing a scale; it turns out that the two parameters of summability are closely connected. The results permit interesting applications to Fourier series and other developments. The last section gives some applications to simple Fourier series. Further applications, in particular, to multiple Fourier series will be presented in another paper.

1. Let $J_\mu(t)$ denote the Bessel function of order μ:

$$J_\mu(t) = t^\mu/2^\mu \sum_{\nu=0}^{\infty} (-1)^\nu (t^{2\nu}/2^{2\nu}\nu!\Gamma(\mu+\nu+1)), \qquad R(\mu) > -\tfrac{1}{2},$$

and let

$$(1.1) \qquad \alpha_\mu(t) = (2^\mu\Gamma(\mu+1)J_\mu(t)/t^\mu)$$
$$= \Gamma(\mu+1)\sum_{\nu=0}^{\infty}(-1)^\nu(t^{2\nu}/4^\nu\nu!\Gamma(\mu+\nu+1)$$

so that $\alpha_\mu(0) = 1$.

Let $\{\lambda_n\}$ be an increasing sequence of positive numbers

$$0 \leq \lambda_0 < \lambda_1 < \lambda_2 < \cdots < \lambda_n \to \infty,$$

and let k denote a positive integer; a series $\sum_{n=0}^{\infty} a_n$ is said to be summable (J_μ, k, λ) to the sum s if the series

$$(1.2) \qquad \sum_{n=0}^{\infty} a_n\{\alpha_\mu(\lambda_n t)\}^k = \phi_\mu^k(t)$$

converges in some interval $0 < t < t_0$, and if

$$\phi_\mu^k(t) \to s \text{ as } t \to 0. \qquad ([1], [2])\,[2]$$

Our transform includes Riemann's summability-methods for $\lambda_n = n$, $\mu = \frac{1}{2}$, $\alpha_{\frac{1}{2}}(t) = \sin t/t$. For $\lambda_n = n$ and $k = 1$ we call it J_μ summability.

* Received August 29, 1947.

[1] Presented to the American Mathematical Society, April 26, 1947.

[2] Numbers in brackets refer to the literature at the end of the paper.

We now develop a number of properties of $\alpha_\mu(t)$ which will be needed in what follows.

We have, for $R(\mu) > -\frac{1}{2}$, ([8], pp. 47, 48)

$$J_\mu(t) = (t/2)^\mu(1/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2}))\int_0^\pi \cos(t\cos\theta)\sin^{2\mu}\theta d\theta,$$

thus

(1.3) $$\alpha_\mu(t) = (\Gamma(\mu+1)/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2})\int_0^\pi \cos(t\cos\theta)\sin^{2\mu}\theta d\theta.$$

It follows that for real t and real $\mu > -\frac{1}{2}$

$$|\alpha_\mu(t)| \leq (\Gamma(\mu+1)/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2}))\int_0^\pi \sin^{2\mu}\theta d\theta = \alpha_\mu(0) = 1,$$

and

$$1-\alpha_\mu(t) = (\Gamma(\mu+1)/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2}))\int_0^\pi \{1-\cos(t\cos\theta)\}\sin^{2\mu}\theta d\theta.$$

Now

$$1-\cos(t\cos\theta) < \tfrac{1}{2}(t\cos\theta)^2,$$

hence

$$0 \leq 1-\alpha_\mu(t) < (\Gamma(\mu+1)/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2})\cdot t^2/2\int_0^\pi \cos^2\theta\sin^{2\mu}\theta d\theta < \tfrac{1}{2}t^2$$

for $\mu > -\frac{1}{2}$.

It is known ([8], p. 199) that

$$J_\mu(t) = (2/\pi t)^{1/2}\cos(t-\mu\pi/2-\pi/4) + O(t^{-3/2}), \text{ as } t\to+\infty,$$

and that ([8], p. 45)

$$(d/dt)(t^{-\mu}J_\mu(t)) = -t^{-\mu}J_{\mu+1}(t);$$

hence

(1.4) $$\alpha_\mu(t) = (2/\pi)^{1/2}(2^\mu\Gamma(\mu+1)/t^{\mu+\frac{1}{2}})\cos(t-\mu\pi/2-\pi/4) + O(t^{-\mu-3/2}) = O(t^{-\mu-\frac{1}{2}}),$$

and

(1.5) $$(d/dt)[\alpha_\mu(t)] = -t\alpha_{\mu+1}(t)/2(\mu+1).$$

All roots of $J_\mu(t)$ are real for $\mu \geq -\frac{1}{2}$ ([8], p. 482), and if t_ν are the positive roots in increasing order, then $t_\nu = O(\nu)$ ([8], p. 506). From (1.3)

$$\alpha'_\mu(t) = (-\Gamma(\mu+1)/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2}))\int_0^\pi \sin(t\cos\theta)\cos\theta\cdot\sin^{2\mu}\theta d\theta,$$

and from the mean-value theorem, for $0 < \tau < t$,

$$\alpha_\mu(t) - \alpha_\mu(\tau) = \alpha'_\mu(x)(t-\tau), \text{ where } \tau < x < t.$$

Hence

(1.6) $|\alpha_\mu(t)-\alpha_\mu\tau)|$

$$\leqq ((t-\tau)\Gamma(\mu+1)x/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2}))\int_0^\pi \cos^2\theta\sin^{2\mu}\theta d\theta < t(t-\tau);$$

in particular

(1.7) $$|\alpha_\mu(nt)-\alpha_\mu(\overline{n+1}\,t)| < (1+n)t^2, \qquad t>0.$$

From (1.4) and (1.5)

$$\alpha_\mu(nt)-\alpha_\mu(\overline{n+1}\,t)$$

(1.8) $$=1/2(\mu+1)\int_{nt}^{(n+1)t} x\alpha_{\mu+1}(x)\,dx = O(\int_{nt}^{(n+1)t} x^{-\mu-\frac{1}{2}}dx)$$
$$=O(t^{\frac{1}{2}-\mu}n^{-\frac{1}{2}-\mu}), \text{ as } n\to\infty.$$

2. THEOREM 1. *If* $0<\alpha<\delta<1$, $S_n=\sum_0^n a_\nu$,

(2.1) $$a_n=O(n^{-\delta}),$$

(2.2) $$S_n-s=o(n^{-\alpha}), \text{ as } n\to\infty,$$

then Σa_n *is summable* J_μ *to the value* s *for* $\mu=\frac{1}{2}-\alpha/1-\delta+\alpha$.

For the proof we may assume $s=0$; let

(2.3) $$r=1/(\delta-\alpha), \text{ so that } r>1/\delta>1.$$

We write for a given ϵ $(o<\epsilon<1)$

(2.4) $$\lambda=[t^{-1}],\ \nu=1+[\epsilon^{-1}t^{-r}], \text{ where } t<1,$$

so that $\lambda\geqq 1$, $\nu>\lambda$. Let

(2.5) $$\phi_\mu(t)=\sum_{n=0}^{\nu} a_n\alpha_\mu(nt)+\sum_{n=\nu+1}^{\infty} a_n\alpha_\mu(nt)=\psi_1(t)+\psi_2(t), \text{ say.}$$

Now, from (2.1) and (1.4)

$$\psi_2(t)=O(\sum_{n=\nu+1}^{\infty} n^{-\delta}(nt)^{-\mu-\frac{1}{2}}=O(t^{-\mu-\frac{1}{2}}\nu^{-\delta-\mu+\frac{1}{2}}),$$

as

$$\mu+\tfrac{1}{2}+\delta=1+\delta-(\alpha/1-\delta+\alpha)=(1-\delta(\delta-\alpha)/1-(\delta-\alpha))>1.$$

This proves that the series $\phi_\mu(t)$ is convergent for $t>0$, and it follows from (2.4) that

$$\psi_2(t)=O(t^{-\mu-\frac{1}{2}+r(\mu+\delta-\frac{1}{2})}\epsilon^{\mu+\delta-\frac{1}{2}}).$$

Here

$$\mu + \delta - \tfrac{1}{2} = \delta - (\alpha/1 - \delta + \alpha) = ((1-\delta)(\delta - \alpha)/1 - \delta + \alpha)) > 0,$$

and the exponent of t is, in view of (2.3)

$$(1 - \delta/1 - \delta + \alpha) - (1 - (\alpha/1 - \delta + \alpha)) = 0.$$

Thus $\psi_2(t)$ is small for small ϵ;

(2.6) $$|\psi_2(t)| < \epsilon', \text{ where } \epsilon' \to o \text{ as } \epsilon \to o.$$

Furthermore

$$\begin{aligned}\psi_1(t) &= \sum_0^{\nu-1} S_n[\alpha_\mu(nt) - \alpha_\mu(\overline{n+1}\,t)] + S_\nu\alpha_\mu(\nu t)\\ &= \psi_3(t) + S_\nu\alpha_\mu(\nu t), \text{ say.}\end{aligned}$$

Now, from (2.2) and (1.4)

$$\begin{aligned}S_\nu\alpha_\mu(\nu t) &= o(\nu^{-\alpha-\frac{1}{2}-\mu}t^{-\frac{1}{2}-\mu})\\ &= \epsilon^{\mu+\frac{1}{2}+\alpha}\cdot o(t^{r(\mu+\frac{1}{2}+\alpha)-\frac{1}{2}-\mu}), \text{ (from (2.4)).}\end{aligned}$$

Here

$$\mu + \tfrac{1}{2} + \alpha = 1 - (\alpha/1 - \delta + \alpha) + \alpha > \alpha > o,$$

and

$$r(\mu + \tfrac{1}{2} + \alpha) - \tfrac{1}{2} - \mu > o;$$

hence

(2.7) $$S_\nu\alpha_\mu(\nu t) = o(1), \text{ as } t \to o.$$

Finally for $\psi_3(t)$ write

$$\psi_3(t) = \Big(\sum_{n=0}^{\lambda} + \sum_{n=\lambda+1}^{\nu-1}\Big) S_n\{\alpha_\mu(nt) - \alpha_\mu(\overline{n+1}\,t)\} = \psi_4(t) + \psi_5(t),$$

say; using (2.2) and (1.7), and noting that (1.7) holds also for $n = 0$ in view of the inequality $1 - \alpha_\mu(t) < \frac{1}{2}t^2$, we obtain

(2.8) $$\psi_4(t) = O\Big(\sum_{n=0}^{\lambda} n^{-\alpha+1}t^2\Big) = O(t^2\lambda^{2-\alpha}) = O(t^\alpha), \text{ (from (2.4)).}$$

To estimate $\psi_5(t)$ we employ (1.8) and (2.2); then

$$\psi_5(t) = o\Big(\sum_{n=\lambda}^{\nu} n^{-\alpha-\frac{1}{2}-\mu}t^{\frac{1}{2}-\mu}\Big) - o(t^{\frac{1}{2}-\mu}\lambda^{\frac{1}{2}-\alpha-\mu}),$$

as

$$\tfrac{1}{2} - \alpha - \mu = (\alpha/1 - \delta + \alpha) - \alpha = (\alpha(\delta - \alpha)/1 - \delta + \alpha)) > 0.$$

Thus, from (2.4)

$$\begin{aligned}\psi_5(t) &= o(t^{\frac{1}{2}-\mu}\epsilon^{\mu-\frac{1}{2}+\alpha}t^{r(\alpha+\mu-\frac{1}{2})})\\ &= \epsilon^{\mu+\alpha-\frac{1}{2}}\cdot o(t^{2(\alpha+\mu-\frac{1}{2})+\frac{1}{2}-\mu}).\end{aligned}$$

The exponent of t is, in view of (2. 3),

$$(-\alpha/1-\delta+\alpha)+(\alpha/1-\delta+\alpha)=o;$$

thus

(2. 9) $$\psi_5(t)=\epsilon^{\alpha+\mu-\frac{1}{2}}\cdot o(1);$$

here the first factor is large for small ϵ, but for any given $\epsilon>o$, choosing t small enough: $t<t_0(\epsilon)$, ψ_5 will become arbitrarily small. It now follows from (2. 5)-(2. 9) that $\limsup_{t\to 0}|\phi_\mu(t)|<\epsilon'$; ϵ' being arbitrarily small, the theorem follows.

For $\alpha=o$ ([7], p. 389) at least one of the two assumptions (2. 1), (2. 2) must be strengthened. In this case sufficient assumptions are, either (2. 1) and

(2. 2′) $$S_n-s=o(1/\log n),$$

or, (2. 1) for *every* $\delta<1$, and

(2. 2″) $$S_n-s=O(1/\log n),$$

or, (2. 1) for $\delta=1$ and (2. 2) (for $\alpha=o$).

See [6] also for further references and generalizations in the case $\alpha=o$. The main results in this case are due to Hardy and Littlewood.

3. Chandrasekharan ([1], Theorem I) proved that summability (C,r), $r\geqq o$ implies summability J_μ for $\mu>r+\frac{1}{2}$ to the same sum. We shall prove the following generalization:

THEOREM 2. *If Σa_n is $(C,r+1)$ summable, and if*

(3. 1) $$\sigma_n=\sum_{\nu=0}^{n}|S_\nu^r|=O(n^{r+1}),\text{ as } n\to\infty,$$

then Σa_n is J_μ summable for $\mu>r+\frac{1}{2}$, $r>-1$.

Here S_n^r is the n-th Cesaro mean of order r:

(3. 2) $$S_n^r=\sum_{\nu=0}^{n}\gamma^r{}_{n-\nu}a_\nu,\quad \gamma_n^r=\binom{n+r}{n}.$$

Let

(3. 3) $$\sum_{n=\nu}^{\infty}\gamma_{n-\nu}{}^{-r-2}\alpha_\mu(nt)=\Delta^{r+1}\alpha_\mu(\nu t),\qquad \nu=0,1,2,\cdots.$$

It is well-known that

$$\gamma_n^r = (r+1)(r+2)\cdots(r+n)/n! \sim n^r/\Gamma(r+1), \tag{3.4}$$

if r is not a negative integer.

For the proof of the theorem we may assume without loss of generality that $(C, r+1)\Sigma a_n = 0$, that is, $S_n^{r+1} = o(n^{r+1})$ as $n \to \infty$. The main idea of the proof is then as follows. Since we have $a_n = \sum_{\nu=0}^{n} \gamma_{n-\nu}^{-r-2} S_\nu^r$, we write (purely formally, for the present)

$$\begin{aligned} \phi_\mu(t) &= \sum_{n=0}^{\infty} \alpha_\mu(nt) \Big(\sum_{\nu=0}^{n} \gamma_{n-\nu}^{-r-2} S_\nu^r \Big) = \sum_{\nu=0}^{\infty} \sum_{n=\nu}^{\infty} S_\nu^r \gamma_{n-\nu}^{-r-2} \alpha_\mu(nt), \\ &= \sum_{\nu=0}^{\infty} S_\nu^r \Delta^{r+1} \alpha_\mu(\nu t), \end{aligned} \tag{3.5}$$

and we then estimate $\Delta^{r+1}\alpha_\mu(\nu t)$ in different ways in different parts of the range of ν in (3.5) in order to show that $\phi_\mu(t) \to 0$.

We first show that the interchange of summation in (3.5) is legitimate by proving that the double series converges absolutely. We have for $t > o$,

$$\begin{aligned} \sum_{n=\nu}^{\infty} | \gamma_{n-\nu}^{-r-2} \alpha_\mu(nt) | &= O\Big(\sum_{n=\nu}^{\infty} (n-\nu+1)^{-r-2} n^{-\mu-\frac{1}{2}} \Big) \\ &= O(\nu^{-\mu-\frac{1}{2}}), \end{aligned}$$

and we need only prove that $\sum_{\nu=1}^{\infty} | S_\nu^r | \nu^{-\mu-\frac{1}{2}} < \infty$. Now

$$\begin{aligned} \sum_{\nu=1}^{n} | S_\nu^r | \nu^{-\mu-\frac{1}{2}} &= n^{-\mu-\frac{1}{2}} \sigma_n + \sum_{1}^{n-1} \sigma_\nu [\nu^{-\mu-\frac{1}{2}} - \overline{\nu+1}^{-\mu-\frac{1}{2}}], \\ &= O(n^{-\mu-\frac{1}{2}+r+1}) + O\Big(\sum_{\nu=1}^{n-1} \nu^{r+1-\mu-3/2} \Big), \end{aligned}$$

from (3.1). Hence

$$\sum_{\nu=1}^{n} | S_\nu^r | \nu^{-\mu-\frac{1}{2}} = O(1) + O(1) = O(1).$$

This proves (3.5) and the convergence of either series.

It remains for us to prove that $\lim_{t\to 0} \phi_\mu(t) = 0$, $\mu > r + \frac{1}{2}$, $r > -1$. Let us write, for any $\lambda > 1$,

$$\phi_\mu(t) = \Big(\sum_{\nu t \leq \lambda} + \sum_{\nu t > \lambda} \Big) S_\nu^r \Delta^{r+1} \alpha_\mu(\nu t) = \psi_1(t) + \psi_2(t), \text{ say.}$$

We shall show that $\psi_1(t)$, $\psi_2(t)$ tend to zero with t. Before doing so, we shall evaluate $\Delta^{r+1}\alpha_\mu(\nu t)$. Employing the formula ([8], p. 48)

$$J_\mu(t)=t^\mu/2^\mu\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2})\int_0^\pi e^{it\cos\theta}\sin^{2\mu}\theta d\theta, \tag{3.6}$$

we get, from (1.1) and (3.3);

$$\begin{aligned}\Delta^{r+1}\alpha_\mu(\nu t)&=\sum_{n=\nu}^{\infty}\gamma_{n-\nu}^{-r-2}(\Gamma(\mu+1)/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2}))\int_0^\pi e^{int\cos\theta}\sin^{2\mu}\theta d\theta\\&=(\Gamma(\mu+1)/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2}))\int_0^\pi\sin^{2\mu}\theta(\sum_{n=\nu}^{\infty}\gamma_{n-\nu}^{-r-2}e^{int\cos\theta})d\theta.\end{aligned}$$

Now

$$\begin{aligned}\Delta^{r+1}e^{i\nu t\cos\theta}&=\sum_{n=\nu}^{\infty}\gamma_{n-\nu}^{-r-2}e^{int\cos\theta}=e^{i\nu t\cos\theta}\cdot\sum_{n=0}^{\infty}\gamma_n^{-r-2}e^{int\cos\theta}\\&=z^\nu(1-z)^{r+1},\end{aligned}$$

where $z=e^{it\cos\theta}$, and the substitution $\cos\theta=x$ yields

$$\Delta^{r+1}\alpha_\mu(\nu t)=C_\mu\int_{-1}^{+1}(1-x^2)^{\mu-\frac{1}{2}}e^{i\nu tx}(1-e^{itx})^{r+1}dx, \tag{3.7}$$

where

$$C_\mu\equiv(\Gamma(\mu+1)/\Gamma(\mu+\tfrac{1}{2})\Gamma(\tfrac{1}{2}))=\left(\int_0^\pi\sin^{2\mu}\theta d\theta\right)^{-1}=\left(\int_{-1}^{+1}(1-x^2)^{\mu-\frac{1}{2}}dx\right)^{-1}.$$

To estimate ψ_1 let $[\lambda t^{-1}]=n$, so that

$$\psi_1(t)=\sum_{\nu=0}^{n}S_\nu^r\Delta^{r+1}\alpha_\mu(\nu t)=\sum_{\nu=0}^{n}(S_\nu^{r+1}-S_{\nu-1}^{r+1})\Delta^{r+1}\alpha_\mu(\nu t),\ S_{-1}^{r+1}=0;$$

then

$$\psi_1(t)=S_n^{r+1}\Delta^{r+1}\alpha_\mu(nt)+\sum_{\nu=0}^{n-1}S_\nu^{r+1}\Delta^{r+2}\alpha_\mu(\nu t). \tag{3.8}$$

Using (3.7) we obtain,

$$\begin{aligned}|\Delta^{r+1}\alpha_\mu(nt)|&<C_\mu\int_{-1}^{+1}(1-x^2)^{\mu-\frac{1}{2}}|1-e^{itx}|^{r+1}dx\\&=C_\mu\int_{-1}^{+1}(1-x^2)^{\mu-\frac{1}{2}}|2\sin tx/2|^{r+1}dx\\&<C_\mu\int_{-1}^{+1}(1-x^2)^{\mu-\frac{1}{2}}t^{r+1}dx=t^{r+1},\end{aligned}$$

so that

$$S_n^{r+1}\Delta^{r+1}\alpha_\mu(nt)=o((nt)^{r+1}).$$

To estimate the second term in (3.8), we note that $|S_\nu^{r+1}|=\epsilon_\nu\nu^{r+1}$ where $\epsilon_\nu\to o$ and hence,

$$\Big|\sum_{\nu=0}^{n-1} S_\nu^{r+1}\Delta^{r+2}\alpha_\mu(\nu t)\Big| \leq t^{r+2}\sum_{\nu=0}^{[\lambda/t]} |S_\nu^{r+1}|$$
$$\leq \lambda\cdot t^{r+1}\cdot(\lambda/t)^{r+1}\cdot t/\lambda\sum_{\nu=0}^{[\lambda/t]}\epsilon_\nu = o(1),$$

by the $(C,1)$ consistency-theorem. Hence it follows that

$$\psi_1(t) = o(1), \text{ as } t\to o.$$

To estimate $\psi_2(t)$ write

$$\Delta^{r+1}\alpha_\mu(\nu t) = \Delta^{r+1-k}\cdot\Delta^k\alpha_\mu(\nu t), \qquad k=[r+1].$$

Using the well-known mean-value theorem for repeated differences,

$$\Delta^\rho\alpha_\mu(\nu t) = (-1)^\rho t^\rho\alpha_\mu^{(\rho)}(\overline{\nu+\delta\,\rho t}),\ o<\delta<1, \qquad \rho=0,1,\cdots.$$

From (1. 4) and (1. 5) we have

$$\alpha'_\mu(t) = O(t^{-\mu-\frac{1}{2}}), \text{ as } t\to\infty,$$

and by induction,

$$\alpha_\mu^{(\rho)}(t) = O(t^{-\mu-\frac{1}{2}}), \qquad \rho=0,1,\cdots;\ t\to\infty.$$

It follows that

$$\Delta^\rho\alpha_\mu(\nu t) = O(t^{\rho-\mu-\frac{1}{2}}\nu^{-\mu-\frac{1}{2}}).$$

First let r be an integer. Then

$$\sum_{\nu=n+1}^{\infty} S_\nu^r\Delta^{r+1}\alpha_\mu(\nu t) = O\Big(\sum_{\nu=n+1}^{\infty}|S_\nu^r|\,t^{r+\frac{1}{2}-\mu}\nu^{-\mu-\frac{1}{2}}\Big)$$
$$= O\Big(t^{r+\frac{1}{2}-\mu}\sum_{\nu=n+1}^{\infty}|S_\nu^r|\,\nu^{-\mu-\frac{1}{2}}\Big)$$
$$= O\Big(t^{r+\frac{1}{2}-\mu}\sum_{\nu=n+1}^{\infty}(\sigma_\nu-\sigma_{\nu-1})\nu^{-\mu-\frac{1}{2}}\Big).$$

Thus

$$\psi_2(\mathrm{t}) = O\Big(t^{r+\frac{1}{2}-\mu}\sum_{\nu=n}^{\infty}\sigma_\nu\nu^{-\mu-3/2}\Big) = O\Big(t^{r+\frac{1}{2}-\mu}\sum_{\nu=n}^{\infty}\nu^{-\mu-\frac{1}{2}+r}\Big)$$
$$= O(t^{r+\frac{1}{2}-\mu}n^{r-\mu+\frac{1}{2}}) = O(\lambda^{r+\frac{1}{2}-\mu}) < \epsilon$$

if $\lambda=\lambda(\epsilon)$ is large enough. Next, let r be non-integral, $r<k<r+1$. Then

$$\Delta^{r+1}\alpha_\mu(\nu t) = \sum_{m=\nu}^{\infty}\gamma_{m-\nu}{}^{k-r-1}\Delta^{k+1}\alpha_\mu(mt) = \Delta^{k+1}\alpha_\mu(\nu t) + \sum_{m=\nu+1}^{\infty}\gamma_{m-\nu}{}^{k-r-1}\Delta^{k+1}\alpha_\mu(\nu t).$$

Therefore we write

$$\psi_2(t) = \psi_3(t)+\psi_4(t)+\psi_5(t),$$

where

$$\psi_3(t) \equiv \sum_{\nu=n+1}^{\infty} S_\nu{}^r \Delta^{k+1}\alpha_\mu(\nu t),$$

$$\psi_4(t) \equiv \sum_{\nu=n+1}^{\infty} S_\nu{}^r \sum_{m=\nu+1}^{\nu+\rho} \gamma_{m-\nu}{}^{k-r-1}\Delta^{k+1}\alpha_\mu(mt),$$

$$\psi_5(t) \equiv \sum_{\nu=n+1}^{\infty} S_\nu{}^r \sum_{m=\nu+\rho+1}^{\infty} \gamma_{m-\nu}{}^{k-r-1}\Delta^{k+1}\alpha_\mu(mt).$$

To estimate $\psi_3(t)$, we observe that

$$\Delta^{k+1}\alpha_\mu(\nu t) = O(t^{k-\mu+\frac{1}{2}}\nu^{-\mu-\frac{1}{2}}),$$

and so

$$\psi_3(t) = O(t^{k-\mu+\frac{1}{2}} \sum_{\nu=n+1}^{\infty} | S_\nu{}^r | \nu^{-\mu-\frac{1}{2}})$$

(3. 9)
$$= O(t^{k+\frac{1}{2}-\mu}n^{r-\mu+\frac{1}{2}}) = O(t^{k-r}\lambda^{r+\frac{1}{2}-\mu})$$
$$= o(1).$$

To estimate $\psi_4(t)$ we use (3. 4) and observe that

$$\sum_{m=\nu+1}^{\nu+\rho} \gamma_{m-\nu}{}^{k-r-1}\Delta^{k+1}\alpha_\mu(mt) = O(t^{k+\frac{1}{2}-\mu} \sum_{m=\nu+1}^{\nu+\rho} (m-\nu)^{k-r-1}m^{-\mu-\frac{1}{2}})$$
$$= O(t^{k+\frac{1}{2}-\mu}\nu^{-\mu-\frac{1}{2}}\rho^{k-r}),$$

and so

$$\psi_4(t) = O(\sum_{\nu=n+1}^{\infty} | S_\nu{}^r | \nu^{-\mu-\frac{1}{2}}t^{k+\frac{1}{2}-\mu}\rho^{k-r})$$

(3. 10)
$$= O(t^{k+\frac{1}{2}-\mu}\rho^{k-r}n^{r-\mu+\frac{1}{2}})$$
$$= O((nt)^{r-\mu+\frac{1}{2}}t^{k-r}\rho^{k-r})$$
$$= O(\lambda^{r+\frac{1}{2}-\mu})(\rho t)^{k-r}.$$

Finally, to estimate $\psi_5(t)$ we observe that

$$\sum_{m=\nu+\rho+1}^{\infty} \gamma_{m-\nu}{}^{k-r-1}\Delta^{k+1}\alpha_\mu(mt) = \gamma_{\rho+1}{}^{k-r-1}\Delta^{k+1}\alpha_\mu(\overline{\nu+\rho+1}\,t) + \sum_{m=\nu+\rho+2}^{\infty} \gamma_{m-\nu}{}^{k-r-2}\Delta^{k}\alpha_\mu(mt)$$
$$= O(\rho^{k-r-1}t^{k-\mu-\frac{1}{2}}(\nu+\rho)^{-\mu-\frac{1}{2}}),$$

and on

$$\psi_5(t) = O(\sum_{\nu=n+1}^{\infty} | S_\nu{}^r | \rho^{k-r-1}t^{k-\mu-\frac{1}{2}}(\nu+\rho)^{-\mu-\frac{1}{2}})$$

(3. 11)
$$= O(n^{r-\mu+\frac{1}{2}}\rho^{k-r-1}t^{k-\mu-\frac{1}{2}})$$
$$= O((nt)^{r-\mu+\frac{1}{2}}t^{k-r-1}\rho^{k-r-1}).$$

Let $\rho = [t^{-1}]$, then from (3.9), (3.10) and (3.11) we have

$$\psi_2(t) = \psi_3 + \psi_4 + \psi_5 = o(1) + O(\lambda^{r+\frac{1}{2}-\mu}) < \epsilon$$

for $\lambda > \lambda(\epsilon)$. It follows that

$$\phi_\mu(t) = \psi_1(t) + \psi_2(t) = o(1), \text{ as } t \to o.$$

Remark. For $r = -\alpha$, $0 < \alpha < 1$, $\mu = \frac{1}{2}$, see Theorem A in [7]. Note that there are series satisfying the assumptions of our theorem but not summable (C, k) for any $k < r + 1$. Cf. [7].

4. In the present section, we assume that Σa_n is summable (C, r) for some $r \geqq 0$. It then follows that Σa_n is J_μ summable for $\mu > r + \frac{1}{2}$, but it is *not* necessarily summable for $\mu = r + \frac{1}{2}$. ([1], p. 226). We shall now investigate conditions under which (C, r) summability of Σa_n will imply $J_{r+\frac{1}{2}}$ summability. In doing so, we will find it convenient to use, instead of Cesaro summability, an *equivalent* type of Riesz summability.

Let us assume that $(C, r)\Sigma a_n = 0$. Let us write

$$T(w) = T^0(w) = \sum_{\nu < w} a_\nu, \qquad \nu \leq w < \nu + 1.$$

In the notation of Sections 2 and 3, $T(w) = S_\nu$. For $k > 0$, we write

$$T^k(w) = 2k \int_0^w (w^2 - t^2)^{k-1} t T(t)\, dt.$$

A series Σa_n is summable (ν^2, k) to the sum s, if $\lim_{w\to\infty} w^{-2k} T^k(w) = s$. It has been shown by Hardy [3] that this method is equivalent to (C, k); thus

$$T^r(w) = o(w^{2r}). \tag{4.1}$$

We also have ([4], p. 36),

$$T(w) = o(w^r), \tag{4.2}$$

and for $k \geq 0$, $\mu > 0$ ([4], p. 27),

$$T^{k+\mu}(w) = (2\Gamma(\mu + k + 1)/\Gamma(k+1)\Gamma(\mu)) \int_0^w (w^2 - t^2)^{\mu-1} t T^k(t)\, dt. \tag{4.3}$$

In the notation of Hardy and Riesz ([4], p. 21)

$$T^k(w) = C_\lambda{}^k(w^2) = k \int_0^{w^2} C_\lambda(t)\,(w^2 - t)^{k-1} dt$$

$$T(w) = C_\lambda(w^2), \qquad \{\lambda_n\} = \{n^2\}.$$

We have ([4], p. 36)

$$T^k(w) = o(w^{r+k}), \text{ for } o \leq k < r. \tag{4.4}$$

We shall now prove some lemmas leading to the main theorem of this section.

LEMMA 1. *If Σa_n is summable (C, r) for some $r \geq o$, to the sum zero, r not an integer, $h = [r]$, $\mu \geq h + \frac{1}{2}$, then we have:*

$$\sum_{n=0}^{m} a_n \alpha_\mu(nt) = o(1) + ct^{2r+2} \int_0^m u T^r(u) \alpha_{\mu+r+1}(ut)\, du;$$

if r is an integer, then the above relation holds for $\mu \geq r - \frac{1}{2}$; if r is zero, it holds for $\mu > -\frac{1}{2}$.

To prove the lemma, we observe that Abel's lemma on partial summation and (1.8) yield

$$\sum_{n=0}^{m} a_n \alpha_\mu(nt) = S_m \alpha_\mu(mt) + t^2/2(\mu+1) \int_0^m x T(x) \alpha_{\mu+1}(xt)\, dx. \tag{4.5}$$

Denote by h the greatest integer less than r, so that $r - 1 \leq h < r$; then (1.4) and (4.4) yield the relations

$$T^k(w) \alpha_{\mu+k}(wt) = o(w^{r+k-\mu-k-\frac{1}{2}}) = o(w^{r-\mu-\frac{1}{2}}) = o(1), \tag{4.6}$$

for $0 < k \leq r$, $\mu \geq r - \frac{1}{2}$.

If r is not an integer, $h + 1 > r$, and since Cesaro summability is regular, the given series is also summable $(C, r+1)$, or what is the same, summable $(\nu^2, r+1)$. Hence

$$T^{h+1}(w) = o(w^{2h+2}), \tag{4.7}$$

and

$$T^{h+1}(w) \alpha_{\mu+h+1}(wt) = o(w^{2h+2} \cdot w^{-(\mu+h+1)-\frac{1}{2}}) = o(1), \text{ if } \mu \geq h + \tfrac{1}{2}. \tag{4.8}$$

If r is an integer, then $h = r - 1$; and

$$T^{h+1}(w) \alpha_{\mu+h+1}(wt) = T^r(w) \alpha_{\mu+r}(wt) = o(1), \text{ if } \mu \geq r - \tfrac{1}{2}$$

by (4.6).

If $r = 0$, hence $h = -1$, then

$$T(w) \alpha_\mu(wt) = o(w^r w^{-\mu-\frac{1}{2}}) = o(w^{-\mu-\frac{1}{2}}) = o(1),\ \mu > -\tfrac{1}{2}.$$

Now, integrating by parts $h + 1$ times in (4.5) and using (4.6), (4.8), we get

$$\sum_{n=0}^{m} a_n \alpha_\mu(nt) = o(1) + ct^{2h+4} \int_0^m x T^{h+1}(x) \alpha_{\mu+h+2}(xt)\, dx, \tag{4.9}$$

where $\mu \geq r - \frac{1}{2}$ if r is an integer > 0; $\mu > -\frac{1}{2}$ if $r = 0$; and $\mu \geq [r] + \frac{1}{2}$ if r is not an integer. The c's stand for constants.

(4. 9), as it is, proves the required result, if r is an integer or zero. If r is not an integer, then employing (4. 7) in (4. 9) we get

$$\sum_{n=0}^{m} a_n \alpha_\mu(nt) = o(1) + ct^{2h+4} \int_0^m x\alpha_{\mu+h+2}(xt) \int_0^x T^r(u)(x^2 - u^2)^{h-r} u\,dudx$$
$$= o(1) + ct^{2h+4} \int_0^m uT^r(u) \int_u^m x\alpha_{\mu+h+2}(xt)(x^2 - u^2)^{h-r} dxdu.$$

We next show that

$$B_m \equiv \int_0^m uT^r(u) \int_m^\infty x\alpha_{\mu+h+2}(xt)(x^2 - u^2)^{h-r} dxdu \to 0 \text{ as } m \to \infty.$$

We have; if the "max" refers to the range $m' \geq m$,

$$\left| \int_m^\infty (x^2 - u^2)^{h-r} x\alpha_{\mu+h+2}(xt)\,dx \right| \leq (m^2 - u^2)^{h-r} \max \left| \int_m^{m'} x\alpha_{\mu+h+2}(xt)\,dx \right|$$
$$= (m^2 - u^2)^{h-r} \max \left| \int_{mt}^{m't} u\alpha_{\mu+h+2}(u)\,du \right| \cdot 1/t^2$$
$$= t^{-2}(m^2 - u^2)^{h-r} \max \left| \int_{mt}^{m't} \alpha'_{\mu+h+1}(u)\,du \right| \cdot 2/\mu + h + 2$$
$$= (2t^{-2}(m^2 - u^2)^{h-r})/(\mu + h + 2) \cdot \max \cdot |\alpha_{\mu+h+1}(m't) - \alpha_{\mu+h+1}(mt)|$$
$$= O(m^2 - u^2)^{h-r} \cdot O(m^{-\mu-h-3/2}).$$

Thus

$$B_m = O\left(\int_0^m u\,|T^r(u)|\,(m^2 - u^2)^{h-r} \cdot m^{-\mu-h-3/2} du \right)$$
$$= o\left(\int_0^m u^{2r+1}(m^2 - u^2)^{h-r} m^{-\mu-h-3/2} du \right).$$
$$= o(m^{2h+2-\mu-h-3/2}) = O(m^{h-\mu+\frac{1}{2}})$$
$$= o(1), \text{ if } \mu \geq h + \tfrac{1}{2}.$$

It follows that

$$(4.10) \qquad \sum_{n=0}^{m} a_n \alpha_\mu(nt)$$
$$= o(1) + ct^{2h+4} \int_0^m uT^r(u) \int_u^\infty x\alpha_{\mu+h+2}(xt)(x^2 - u^2)^{h-r} dxdu.$$

In trying to simplify the right side, we shall employ the formula ([8], p. 417, formula 5)

$$\int_0^\infty (J_\nu(a\sqrt{t^2 + z^2})/(t^2 - z^2)^{\nu/2}) t^{2\mu+1} dt = 2^\mu \Gamma(\mu + 1)(J_{\nu-\mu-1}(az)/a^{\nu+1} z^{\nu-\mu-1}),$$

where $a > 0$, $R(\nu/2 - 1/4) > R(\mu) > -1$. We shall apply it only for the case $\nu/2 - 1/4 \geq \mu + 1 > 0$.

Replacing $t^2 + z^2$ by τ^2, and μ by ρ, we get,

$$\int_z^\infty (J_\nu(a\tau)/\tau^{\nu-1})(t^2 - z^2)^\rho d\tau = 2^\rho \Gamma(\rho + 1) J_{\nu-\rho-1}(az)/a^{\rho+1}z^{\nu-\rho-1},$$

$a > 0$, $\nu - \frac{1}{2} \geq 2\rho + 2 > 0$. Or, using (1. 1), we get

$$a^\nu \int_z^\infty \tau\alpha_\nu(a\tau)(\tau^2 - z^2)^\rho d\tau = ca^{\nu-2\rho-2}\alpha_{\nu-\rho-1}(az).$$

Replacing z by u, ν by $\mu + h + 2$, ρ by $h - r$, τ by x, a by t, we get,

$$\int_u^\infty x\alpha_{\mu+h+2}(tx)(x^2 - u^2)^{h-r}dx = ct^{2r-2h-2}\alpha_{\mu+r+1}(ut), \tag{4. 11}$$

for $t > 0$, $\mu + h + 3/2 \geq 2(h - r + 1) > 0$.

But $\mu \geq h + \frac{1}{2}$ and $-1 \leq r - 1 < h < r$; hence $\mu - h + 2r \geq \frac{1}{2}$ and $h - r > -1$.

Now we use (4. 11) in (4. 10) and obtain,

$$\begin{aligned}\sum_{n=0}^m a_n\alpha_\mu(nt) &= o(1) + ct^{2h+4}\int_0^m uT^r(u)t^{2r-2h-2}\alpha_{\mu+r+1}(ut)\,du \\ &= o(1) + ct^{2r+2}\int_0^m uT^r(u)\alpha_{\mu+r+1}(ut)\,du\end{aligned} \tag{4. 12}$$

which completes the proof.

We next prove

LEMMA 2. *If for some* $\lambda > 1$,

$$\int_\omega^{\lambda\omega} |\, dT^r(x)| = O(\omega^{2r}), \text{ as } \omega \to \infty,$$

then, for $\tau > 2r > 0$,

$$\int_\omega^\infty x^{-\tau} |\, dT^r(x)| = O(\omega^{2r-\tau}), \text{ as } \omega \to \infty, \tag{4. 13}$$

and

$$\int_0^\omega x^2 |\, dT^r(x)| = O(\omega^{2r-2}). \tag{4. 14}$$

We have

$$\begin{aligned}\int_\omega^\infty x^{-\tau} |\, dT^r(x)| &= \sum_{\nu=0}^\infty \int_{\lambda^\nu\omega}^{\lambda^{\nu+1}\omega} = O(\sum_{\nu=0}^\infty (\lambda^\nu\omega)^{2r-\tau}) \\ &= O(\omega^{2r-\tau}\sum_{\nu=0}^\infty \lambda^{\nu(2r-\tau)}) = O(\omega^{2r-\tau}).\end{aligned}$$

Furthermore,

$$\int_0^\omega x^2\,|\,dT^r(x)\,| = \sum_{\nu=1}^\infty \int_{\lambda^{-\nu}\omega}^{\lambda^{1-\nu}\omega} x^2\,|\,dT^r(x)\,| = \sum_{\nu=1}^\infty O((\lambda^{1-\nu}\omega)^2(\lambda^{-\nu}\omega)^{2r})$$

$$= O(\omega^{2r+2}\sum_{\nu=1}^\infty \lambda^{-2\nu(r+1)}) = O(\omega^{2r+2}).$$

We can now prove the main theorem of this section.

THEOREM 3. *If Σa_n is summable (C, r) for some $r \geqq 0$, and if for some $\lambda > 1$,*

$$\int_\omega^{\lambda\omega} |\,dT^r(x)\,| = O(\omega^{2r}), \text{ as } \omega \to \infty,$$

then Σa_n is summable J_μ where $\mu > r - \frac{1}{2}$ if r is an integer or zero, and $\mu \geq [r] + \frac{1}{2}$, if r is not an integer.

From Lemma 1, integrating by parts, and using (1.5), we get

$$\sum_{n=0}^m a_n\alpha_\mu(nt) = o(1) + [c_1 t^{2r} T^r(u)\alpha_{\mu+r}(ut)]_0^m + c_2 t^{2r}\int_0^m \alpha_{\mu+r}(ut)\,dT^r(u)$$

$$= o(1) + c_2 t^{2r}\int_0^m \alpha_{\mu+r}(ut)\,dT^r(u).$$

We have by Lemma 2,

$$\left|\int_m^\infty \alpha_{\mu+r}(ut)\,dT^r(u)\right| = O\left(\int_m^\infty (ut)^{-\mu-r-\frac{1}{2}}\,|\,dT^r(u)\,|\right)$$

$$= O(t^{-\mu-r-\frac{1}{2}} m^{r-\mu-\frac{1}{2}}),\ \mu > r - \tfrac{1}{2}.$$

It follows that $\sum_{n=0}^\infty a_n\alpha_\mu(nt)$ converges for $t > 0$, and

$$\sum_{n=0}^\infty a_n\alpha_\mu(nt) = c_2 t^{2r}\int_0^\infty \alpha_{\mu+r}(ut)\,dT^r(u).$$

We write

$$I = t^{2r}\int_0^\infty \alpha_{\mu+r}(ut)\,dT^r(u) = t^{2r}\{\int_0^\omega + \int_\omega^{\omega'} + \int_{\omega'}^\infty\}$$

$$= I_1 + I_2 + I_3, \text{ say.}$$

Now

$$I_1 = t^{2r}\int_0^\omega dT^r(x) + t^{2r}\int_0^\omega \{\alpha_{\mu+r}(ut) - 1\}\,dT^r(u).$$

Given $0 < \epsilon < 1$, we choose ω so large that

$$|\,\omega^{-2r}T^r(\omega)\,| < \epsilon^{2r+2}.$$

We then choose $t = \epsilon/\omega$ and $\omega' = 1/\epsilon t = \omega/\epsilon^2 > \omega$. Then,

$$t^{2r} \mid T^r(\omega) \mid = \mid (\epsilon/\omega)^{2r} T^r(\omega) \mid < \epsilon^{4r+2}.$$

Using (1.6) and (4.14), we get

$$\begin{aligned} \mid \int_0^{\omega} \{\alpha_{\mu+r}(ut) - 1\} dT^r(u) \mid &\leq \int_0^{\omega} \mid \alpha_{\mu+r}(ut) - 1 \mid \cdot \mid dT^r(u) \mid \\ &< t^2 \int_0^{\omega} u^2 \mid dT^r(u) \mid = O(t^2 \omega^{2r+2}) \end{aligned}$$

hence

$$\mid I_1 \mid < \epsilon + O(\epsilon) = O(\epsilon). \tag{4.15}$$

Furthermore, from (4.13),

$$\begin{aligned} I_3 &= O(t^{r-\mu-\frac{1}{2}} \int_{\omega'}^{\infty} u^{-\mu-r-\frac{1}{2}} \mid dT^r(u) \mid) = O((t\omega')^{r-\mu-\frac{1}{2}}) \\ &= O(1) \cdot (1/\epsilon)^{r-\mu-\frac{1}{2}} = o(1), \text{ as } \epsilon \to o. \end{aligned} \tag{4.16}$$

To estimate I_2 consider the successive extrema of $\alpha_{\mu+r}(x)$, which are the roots of $J_{\mu+r+1}(x) = 0$. Denoting the roots of $J_n(x)$ in increasing order by x_k, $k = 0, 1, 2, 3, \cdots$ we have ([8], p. 506)

$$x_k = (k + 3/4 + n/2)\pi + O(1/k), \text{ as } k \to \infty. \tag{4.17}$$

Denote the zeros of $J_{\mu+r+1}(x)$ which lie between ωt and $\omega' t$ by $\xi_1, \xi_2, \cdots, \xi_s$; then $\alpha_{\mu+r}(ut)$ will be monotonic in the intervals,

$$\omega < u < \xi_1/t,\ \xi_1/t < u < \xi_2/t, \cdots, \xi_s/t < \omega' < \xi(s+1)/t.$$

We may suppose, without loss of generality, that $\alpha_{\mu+r}$ is decreasing in the first interval. We write

$$\begin{aligned} I_2 &= t^{2r}\{\int_{\omega}^{\xi_1/t} + \int_{\xi_1/t}^{\xi_2/t} + \cdots + \int_{\xi_s/t}^{\omega'}\} \\ &= I_{2,1} + \cdots + I_{2,s+1}. \end{aligned}$$

Now by the second mean-value theorem,

$$I_{2,1} = t^{2r} \alpha_{\mu+r}(\omega t) \int_{\omega}^{\omega^1} dT^r(u),\ \omega < \omega_1 < \xi_1/t$$

so that

$$\begin{aligned} I_{2,1} &= t^{2r} \alpha_{\mu+r}(\omega t) \{T^r(\omega_1) - T^r(\omega)\} = O(t^{2r}(\xi_1/t)^{2r} \epsilon^{2r+2}) \\ &= O(1) \cdot \xi_1^{2r} \cdot \epsilon^{2r+2}. \end{aligned}$$

Similarly

$$I_{2,2}=t^{2r}\alpha_{\mu+r}(\xi_2)\int_{\omega_2}^{\xi_2/t}dT^r(u)\,;\ \xi_1/t<\omega_2<\xi_2/t,$$

so that

$$I_{2,2}=O(1)\cdot\xi_2^{2r}\epsilon^{2r+2}.$$

Estimating thus each part of I_2, we obtain

$$I_2=O(1)\cdot(\epsilon^{2r+2}\sum_{\nu=1}^{s+1}\xi_\nu^{2r}).$$

But from (4.17) $\xi_\nu=O(\omega t+\nu)$, so that

$$\sum_{\nu=1}^{s+1}\xi_\nu^{2r}=O(\sum_{\nu=1}^{s+1}(\omega t+\nu)^{2r})=O(\int_{\omega t}^{\omega' t}x^{2r}dx)=O(t^{2r+1}\{(\omega^{\mathrm{i}})^{2r+1}-\omega^{2r+1}\})$$
$$=O(t^{2r+1}\omega^{2r+1})\cdot((1/\epsilon^{4r+2})-1).$$

It follows that

$$I_2=O(1)\cdot\epsilon^{4r+3}((1/\epsilon^{4r+2})-1)=o(1). \tag{4.18}$$

Finally (4.16)-(4.18) yield $I\to 0$ as $t\to 0$, which proves Theorem 3. For $r=0$ we find the

COROLLARY. *If Σa_n converges, and if $\sum_{\omega}^{\lambda\omega}|a_n|=O(1)$, as $\omega\to\infty$, $\lambda>1$, then Σa_n is summable J_μ for $\mu>-\frac{1}{2}$. For $\mu=\frac{1}{2}$ see* [7].

5. Chandrasekharan ([1], Theorem II) proved that, under certain conditions, summability (C,r) implies summability $J_{r+\frac{1}{2}+\epsilon}$ and conversely J_μ implies $(C,\mu+\frac{1}{2}+\epsilon)$. A natural converse of the former would, however, be $J_\mu\to(C,\mu-\frac{1}{2}+\epsilon)$, but this holds only under special conditions, and the best result known to the authors in this direction is given below.

We start with the following formula ([1], p. 228): for $\nu>\mu+1$

$$\int_0^\infty\frac{J_\nu(\omega t)J_\mu(nt)}{t^{\nu-\mu-1}}\,dt=\begin{cases}\dfrac{n^\mu(1-n^2/\omega^2)^{\nu-\mu-1}}{2^{\nu-\mu-1}\omega^{2\mu+2-\nu}\Gamma(\nu-\mu)}, & 0<n<\omega,\\ 0,\ n\geqq\omega,\end{cases}$$

or,

$$\int_0^\infty\frac{t^\nu\alpha_\nu(\omega t)\omega^\nu}{2^\nu\Gamma(\nu+1)}\,\frac{t^\mu n^\mu\alpha_\mu(nt)}{2^\mu\Gamma(\mu+1)}\,\frac{dt}{t^{\nu-\mu-1}}$$
$$=\int_0^\infty\frac{t^{2\mu+1}\alpha_\nu(\omega t)\alpha_\mu(nt)\omega^\nu n^\mu}{2^{\nu+\mu}\Gamma(\nu+1)\Gamma(\mu+1)}\,dt=\begin{cases}\dfrac{n^\mu(1-n^2/\omega^2)^{\nu-\mu-1}}{2^{\nu-\mu-1}\omega^{2\mu+2-\nu}\Gamma(\nu-\mu)}, & 0<n<\omega,\\ 0,\ n\geqq\omega.\end{cases}$$

Hence

$$(5.1)\qquad \int_0^\infty \frac{t^{2\mu+1}\alpha_\nu(\omega t)\alpha_\mu(nt)\,dt}{\Gamma(\nu+1)\Gamma(\mu+1)} = \begin{cases} \dfrac{2^{2\mu+1}(1-n^2/\omega^2)^{\nu-\mu-1}}{\omega^{2\mu+2}\Gamma(\nu-\mu)}, & 0<n<\omega \\ 0, & n \geqq \omega. \end{cases}$$

and

$$(5.2)\qquad \sum_{n<\omega} a_n(1-n^2/\omega^2)^{\nu-\mu-1} = \frac{\omega^{2\mu+2}\Gamma(\nu-\mu)2^{-2\mu-1}}{\Gamma(\nu+1)\Gamma(\mu+1)}\int_0^\infty t^{2\mu+1}\alpha_\nu(\omega t)\sum_{n<\omega} a_n\alpha_\mu(nt)\,dt.$$

We now assume that

(5.3) $\sum_{n=0}^{\infty} a_n\alpha_\mu(nt)$ converges dominatedly for $t>0$ to $\phi_\mu(t)$, that is,

$$\Big|\sum_{\nu=0}^{n} a_n\alpha_\mu(\nu t)\Big| \leq O(t),\quad \int_0^\infty O(t)\,dt < \infty;$$

(5.4) $\phi_\mu(t)\to 0$ as $t\to 0$;

(5.5) $\psi(t) = \int_0^t |\phi_\mu(x)|\,dx = O(t)$, as $t\to\infty$;

(5.6) $\omega^\delta \int_{1/\omega}^\infty t^{\delta-1}\phi_\mu(t)\cos(\omega t - \overline{2\mu+1-\delta}(\pi/2))\,dt \to 0$, as $\omega\to\infty$.

From (5.1)-(5.3),

$$\begin{aligned} F(\omega) &= \omega^{2\mu+2}\int_0^\infty \alpha_\nu(\omega t)\phi_\mu(t)t^{2\mu+1}dt \\ &= \sum_{n=0}^{\infty} a_n \int_0^\infty \alpha_\nu(\omega t)\alpha_\mu(nt)t^{2\mu+1}\omega^{2\mu+2}dt \\ &= \omega^{2\mu+2}\sum_{n<\omega}\int_0^\infty t^{2\mu+1}\alpha_\nu(\omega t)\alpha_\mu(nt)\,dt \\ &= \frac{\Gamma(\nu+1)\Gamma(\mu+1)}{\Gamma(\nu-\mu)}2^{2\mu+1}\sum_{n<\omega} a_n(1-n^2/\omega^2)^{\nu-\mu-1}. \end{aligned}$$

We write

$$F(\omega) = \omega^{2\mu+2}\Big\{\int_0^{1/\omega} + \int_{1/\omega}^\infty\Big\} = F_1 + F_2,\ \text{say}.$$

Now

$$|F_1| < \omega^{2\mu+2}\int_0^{1/\omega}|\phi_\mu(t)|\,t^{2\mu+1}dt < \omega\int_0^{1/\omega}|\phi_\mu(t)|\,dt \to 0 \quad\text{as}\quad \omega\to\infty,$$

by (5.4). Furthermore using (1.4),

$$\begin{aligned} F_2 &= c\omega^{3\mu+2}\int_{1/\omega}^\infty \phi_\mu(t)t^{2\mu+1}\cos((\omega t - \overline{\nu-\tfrac12})\pi/2)(\omega t)^{-\nu-\frac12}dt \\ &\quad + O(\omega^{2\mu+2-\nu-3/2}\int_{1/\omega}^\infty|\phi_\mu(t)|\,t^{2\mu+1-\nu-3/2}dt) \\ &= c\omega^{2\mu-\nu+3/2}\int_{1/\omega}^\infty \phi_\mu(t)t^{2\mu+\frac12-\nu}\cos((\omega t-\overline{\nu-\tfrac12})\pi/2)\,dt \\ &\quad + O(\omega^{2\mu+\frac12-\nu}\int_{1/\omega}^\infty|\phi_\mu(t)|\,t^{2\mu-\frac12-\nu}dt). \end{aligned}$$

Putting $\nu = 2\mu + 3/2 - \delta$, $0 \leq \delta < 1$ we get,

$$F_2 = c\omega^\delta \int_{1/\omega}^{\infty} \phi_\mu(t) t^{\delta-1} \cos\left(\left(\omega t - \overline{2\mu + 1 - \delta}\right)\pi/2\right) dt$$
$$+ O\left(\omega^{\delta-1} \int_{1/\omega}^{\infty} |\phi_\mu(t)|\, t^{\delta-2} dt\right)$$
$$= o(1) + O\left(\omega^{\delta-1} \int_{1/\omega}^{\infty} t^{\delta-2} d\psi\right)$$
$$= o(1) + O\left(\omega^{\delta-1}\left\{\left[t^{\delta-2}\psi(t)\right]_{1/\omega}^{\infty} + 2\delta \int_{1/\omega}^{\infty} \psi(t) t^{\delta-3} dt\right\}\right)$$
$$= o(1) + O\left(\omega^{\delta-1}\left\{\left[\int_{1/\omega}^{1} + \int_{1}^{\infty}\right]\psi(t) t^{\delta-3} dt\right\}\right).$$

Finally

$$\omega^{\delta-1} \int_1^\infty \psi(t) t^{\delta-3} dt = O\left(\omega^{\delta-1} \int_1^\infty t^{\delta-2} dt\right) = o(1),$$

and

$$\omega^{\delta-1} \int_{1/\omega}^{1} \psi(t) t^{\delta-3} dt = O\left(\omega^{\delta-1} \int_{1/\omega}^{1} t^{\delta-2} dt\right) = o(1), \text{ as } \omega \to \infty.$$

Using the second consistency theorem (see [3]), we have thus proved:

THEOREM 4. *If (5.3)-(5.6) hold, then Σa_n is summable $(C, \mu + \frac{1}{2} - \delta)$ where $0 \leq \delta < 1$, $\mu > \delta - \frac{1}{2}$ to the sum zero.*

Remark. It may be noted that (5.6) is the only extra condition here which was not required in Chandrasekharan's proof of the result that J_μ implies $(C, \mu + \frac{1}{2} + \epsilon)$. See ([1], Theorem II).

6. In this section, it is shown that summability J_μ of the Fourier series of a function $\phi(t)$ is equivalent to the existence of a generalized limit of the function $\phi(t)$; and the latter is equivalent to the Cesaro limit of ϕ. We thereby deduce, from our previous theorems, which are 'arithmetic' in nature, results concerning the mean continuity of a function and the Cesaro summability of its Fourier series.

Let

$$\phi(t) = \phi_0(t) = f(x+t) + f(x-t)/2 = \phi^*_0(t)$$

$$(6.1) \qquad \phi_p(t) = p/t^p \int_0^t (t-u)^{p-1}\phi(u)\,du, \; p > 0$$

$$(6.2) \qquad \phi^*_p(t) = 2\Gamma(p+\tfrac{1}{2})/\Gamma(\tfrac{1}{2})\Gamma(p)\, t^{2p-1} \int_0^t (t^2 - u^2)^{p-1}\phi(u)\,du, \; p > 0.$$

THEOREM 5. *If $\phi^*_p(t)$ tends to a limit l as $t \to 0$, then $\phi_p(t)$ tends to l as $t \to 0$, and conversely.*

We have

$$\phi_p(t) = p/t^p \int_0^t (t^2 - u^2)^{p-1}\phi(u)(t+u)^{1-p}du \tag{6.3}$$

If $\phi(t) \equiv 1$, then,

$$\int_0^t (t^2 - u^2)^{p-1}\phi(u)\,du = \int_0^t (t^2-u^2)^{p-1}du = t^{2p-1}/2\int_0^1 (1-x)^{p-1}x^{-\frac{1}{2}}dx$$
$$= t^{2p-1}\Gamma(p)\Gamma(\tfrac{1}{2})/2\Gamma(p+\tfrac{1}{2}) = t^{2p-1}c_p, \text{ say;}$$

hence $\phi^*_p(t) \equiv 1$, also $\phi_p(t) \equiv 1$. We may therefore assume that $l = 0$.

Let first $0 < p \leq 1$. In this case the proof is simple. Employing the second mean-value theorem in (6.3) yields

$$\begin{aligned} t^p\phi_p(t) &= p(2t)^{1-p}\int_\zeta^t (t^2-u^2)^{p-1}\phi(u)\,du,\ 0 < \zeta < t, \\ &= p(2t)^{1-p}\{(t^{2p-1}/c_p)\phi^*_p(t) - \int_0^\zeta (t^2-u^2)^{p-1}\phi(u)\,du\} \\ &= t^p\cdot o(1) - p(2t)^{1-p}\int_0^\zeta (t^2-u^2)^{p-1}\phi(u)\,du, \end{aligned}$$

by assumption. We now employ the following inequality, due to M. Riesz ([4], p. 28):

$$\left|\int_0^\tau (t-x)^{p-1}f(x)\,dx\right| \leq \max_{0\leq y\leq\tau}\left|\int_0^y (y-x)^{p-1}f(x)\,dx\right|,$$
$$0 \leq \tau \leq t,\ 0 < p \leq 1.$$

A change of variable $u^2 = x$ yields the analogous inequality

$$\left|\int_0^\xi (t^2-u^2)^{p-1}\phi(u)\,du\right| \leq \max_{0\leq v\leq\xi}\left|\int_0^v (v^2-u^2)^{p-1}\phi(u)\,du\right|,$$
$$0 < p \leq 1,\ 0 \leq \xi \leq t.$$

Using again the assumption $\phi^*_p(t) \to 0$, the first part of theorem 5 follows for $p \leq 1$. The second part follows similarly.

For $p > 1$ we have to consider the two cases, p integral and p non-integral, separately, and adopt substantially the same argument as Hardy does in [3]; only, instead of $t \to \infty$ as in his proof, we have $t \to 0$. Since the proof referred to will run to several pages, and will be new only in details, we choose not to reproduce it here.

Let $f(x) \in L$ and periodic with period 2π;

$$f(x) \sim a_0/2 + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx) \equiv \sum_{n=0}^{\infty} A_n(x) = \Sigma A_n.$$

Then

$$\phi(t) \sim \sum_{n=0}^{\infty} A_n \cos nt = \sum_{n=0}^{\infty} A_n \alpha_{-\frac{1}{2}}(nt),$$

and

$$\phi^*_p(t) = t^{1-2p}/c_p \int_0^t (t^2 - u^2)^{p-1} \sum_{n=0}^{\infty} A_n \alpha_{-\frac{1}{2}}(nu)\, du$$

$$= t^{1-2p}/c_p \sum_{n=0}^{\infty} A_n \cdot \int_0^t (t^2 - u^2)^{p-1} \alpha_{-\frac{1}{2}}(nu)\, du.$$

Added in the proofs: It can be shown that term-wise integration is permitted [Cf. L. S. Bosanquet, *Proc. Lond. Math. Soc.*, vol. 31 (1930), p. 156].

A formula of Sonine ([8], p. 372) yields

$$\alpha_\nu(x) = \frac{2\Gamma(\nu+1)}{\Gamma(\mu+1)\Gamma(\nu-\mu)x^{2\mu+2}} \int_0^x (1 - y^2/x^2)^{\nu-\mu-1} y^{2\mu+1} \alpha_\mu(y)\, dy, \; \nu > \mu;$$

hence, for $\mu = -\frac{1}{2}$, $\nu = p - \frac{1}{2}$,

$$\phi^*_p(t) = \sum_{n=0}^{\infty} A_n \alpha_{p-\frac{1}{2}}(nt).$$

We therefore have the following

LEMMA 3. *If $\Sigma A_n \cos nt$ is a Fourier series, then summability J_μ of ΣA_n for some $\mu > -\frac{1}{2}$ to s implies $\phi^*_{\mu+\frac{1}{2}}(t) \to s$ as $t \to 0$, and vice versa.*

With the aid of this lemma and Theorem 5, *each* of our previous theorems yields a theorem for Fourier series. Thus using Theorem 1, we get

THEOREM 6. *If $0 < \alpha < \delta < 1$, $S_n = \sum_{\nu=0}^{n} A_\nu$* and

$$A_n = O(n^{-\delta}), \qquad S_n - s = o(n^{-\alpha}), \text{ as } n \to \infty,$$

then

$$\phi_p(t) \to s \text{ as } t \to 0, \text{ for } p = 1 - (\alpha/1 - \delta + \alpha) = 1 - \delta/1 - \delta + \alpha.$$

Similarly, Theorem 2 yields

THEOREM 7. *If ΣA_n is summable $(C, r+1)$ to s, and if*

$$\sum_{\nu=0}^{n} | S_\nu{}^r | = O(n^{r+1}), \text{ as } n \to \infty,$$

then $\phi_p(t) \to s$ as $t \to 0$, for $p > r + 1$, where $r > -1$.

This includes the well-known theorem of Paley ([5], p. 190) that $(C, r)\Sigma A_n = s$ implies $\phi_p(t) \to s$ for $p > r + 1$. Paley proved (ibid., p. 199) that the conclusion need not be true for $p = r + 1$. We have however replaced Paley's assumption by an actually more general one.

It is now clear that Theorems 3 and 4 can also be similarly interpreted.

THE INSTITUTE FOR ADVANCED STUDY, PRINCETON,
UNIVERSITY OF CINCINNATI.

REFERENCES.

1. K. Chandrasekharan, "Bessel summation of series," *Proceedings of the Indian Academy of Sciences*, vol. 17 (1943), pp. 219-229.
2. S. Minakshisundaram, "A new summation process," *Mathematics Student*, vol. 11 (1943), pp. 21-27.
3. G. H. Hardy, "The second consistency theorem for summable series," *Proceedings of the London Mathematical Society*, (2) vol. 15 (1916), pp. 72-91.
4. G. H. Hardy and M. Riesz, *The general theory of Dirichlet series*, Cambridge, 1915.
5. R. E. A. C. Paley, "On the Cesaro summability of Fourier series and allied series," *Proceedings of the Cambridge Philosophical Society*, vol. 26 (1930), pp. 173-203.
6. O. Szász, "On convergence and summability of trigonometric series, *American Journal of Mathematics*, vol. 64 (1942) pp. 575-591.
7. O. Szász, "On Lebesgue summability and its generalization to integrals," *American Journal of Mathematics*, vol. 67 (1945), pp. 389-396.
8. G. N. Watson, *Theory of Bessel functions*, 2nd edition, Cambridge, 1944.

WATER WAVES OVER A CHANNEL OF FINITE DEPTH WITH A DOCK.*[1]

By Albert E. Heins.

1. Introduction. The linearized theory of water waves in channels of finite depth has been considered by Stokes who in the course of his work found the periodic solutions. An investigation of the non-periodic solutions in this theory was carried out by A. Weinstein, [2]. Various methods have been used to reconsider this last problem. We mention here the work of Hoheisel (1931) [3], Bochner (1932) [4], Poritsky (1938) [5], Cooper (1939) [6], and Heins (1943) [7]. Another group of problems belonging to this general theory has been recently considered by Miche (1944) [8], Lewy (1946) [9], and Stoker (1947 [10]. These three authors have been concerned with the problem of waves on sloping beaches and have made considerable progress with this problem.

We consider here the problem of a dock over a channel of finite depth, and the possible wave-like solutions. This problem may be formulated as a Wiener-Hopf integral equation and as such, there is mathematical machinery available to solve it. Our method of formulation and mode of solution have several points in common with an investigation in a different field by H. Feshbach and the present author [11]. It will be shown elsewhere that some of the problems considered by Miche, Lewy, and Stoker may also be formulated as integral equations with *Faltung* kernels. In particular, some of their problems may be formulated as generalized Picard integral equations.

We shall not discuss the physical background of the problem we treat here, since it has been adequately described by Lewy, Miche, Stoker, and Weinstein. Suffice it to say, we are concerned with the study of surface waves in an infinite channel of finite depth, and the effect of a semi-infinite rigid dock on these waves. This dock is located on the upper surface of the channel. Mathematically, we are concerned with a solution of Laplace's equation over an infinite slab of width a [12]. That is, we are concerned with the solution of the partial differential equation

$$\Delta\Phi = \Phi_{xx} + \Phi_{yy} + \Phi_{zz} = 0 \tag{1.1}$$

* Received September 3, 1947.

[1] Presented to the American Mathematical Society, April 26, 1947.

over this slab. $\Phi(x, y, z)$ is the velocity potential from which we may derive the x, y, and z components of velocity.

For the boundary conditions to be imposed upon equation (1.1), it is necessary to recall the geometry of the infinite slab. Let the edge of the dock be parallel to the z axis of an xyz coordinate system. We shall measure the depth y, perpendicular to the plane of the dock. The z and y axes now determine a plane and perpendicular to this plane we shall measure x. The dock is a plane which is a units from the bottom of the channel, and is defined for all z and $x < 0$. The free surface of the liquid in the channel is defined in this theory in terms of the boundary condition (1.4). The bottom of the channel has the same physical characteristics as the dock and is defined for all x and z. On a rigid surface, that is, the bottom of the channel and the dock, the normal component of the velocity vanishes. We have then that

$$(1.2) \qquad \Phi_y = 0 \qquad y = 0, -\infty < x < \infty, -\infty < z < \infty$$

and

$$(1.3) \qquad \Phi_y = 0 \qquad y = a,\ x < 0,\ -\infty < z < \infty.$$

On the free surface we have

$$(1.4) \qquad \Phi_{tt} + g\Phi_y = 0 \qquad y = a,\ x > 0,\ -\infty < z < \infty,$$

where t is the time variable and has units which are the same as those used in the gravitational constant g. This boundary condition is a consequence of Bernoulli's law and the assumption that the non-linear terms in the displacement and velocity of the free surface can be neglected. If we now assume that the time dependence of $\Phi(x, y, z, t)$ is monochromatic, that is

$$\Phi(x, y, z, t) = \Phi_1(x, y, z) \exp(ift)$$

equation (1.4) reduces to

$$(1.4a) \qquad \Phi_{1y} = \beta\Phi_1 \qquad y = a,\ x > 0,\ -\infty < z < \infty : \beta = f^2/g > 0$$

while equation (1.1) remains unaltered save for the fact that Φ_1 replaces Φ.

While we have apparently set out to solve a three dimensional problem, we shall reduce it to a two dimensional problem by the further assumption that

$$\Phi_1(x, y, z) = \phi(x, y) \exp(ikz)$$

where k is a constant to which we shall presently ascribe a geometrical interpretation. With this assumption, equation (1.1) becomes

(1. 5) $$\phi_{xx} + \phi_{yy} - k^2\phi = 0$$

while the boundary conditions (1. 2), (1. 3) and (1. 4) become

(1. 2a) $$\phi_y = 0 \qquad y = 0, \ -\infty < x < \infty$$

(1. 3a) $$\phi_y = 0 \qquad y = a, \ x < 0$$

(1. 4a) $$\phi_y = \beta\phi \qquad y = a, \ x > 0$$

Now we are confronted with a strictly two dimensional problem and we have two further sets of conditions to impose on the solution of equation (1. 5). One set of conditions deals with the growth of $\phi(x, y)$ at infinity. Under the dock for $x < < 0$, we would expect little disturbance. We may see this by expanding the solution of equation (1. 5) in terms of the characteristic functions of the strip $0 \leq y \leq a$, $x < 0$. These characteristic functions are

$$\exp\left[x\{k^2 + n^2\pi^2/a^2\}^{\frac{1}{2}}\right] \cos (n\pi y)/a \qquad n = 0, 1, \cdots$$

while

$$\phi(x, y) = \sum_{n=0}^{\infty} A_n \exp\left[x\{k^2 + n^2\pi^2/a^2\}^{\frac{1}{2}}\right] \cos (n\pi y)/a.$$

The A_n's are unknown expansion coefficients. For x large and negative, $\phi(x, y) = (e^{|k|x})$, that is, it is an exponentially decreasing function. We note that a complete knowledge of the A_n's provide us with the solution to our problem, but this is very difficult to obtain directly.

For $x > 0$, the appropriate characteristic functions are

$$\exp\left[-x\{k^2 + C_n^2\}^{\frac{1}{2}}\right] \cos C_n y \qquad n = 1, 2, \cdots$$

where the C_n are the real roots of the transcendental equation

(1. 6) $$\beta \cos Ca + C \sin Ca = 0.$$

There are also two imaginary roots for the equation (1. 6) which we shall denote by $\pm iC_0$. This gives us the possibility of two more characteristic functions

$$\exp\left[\pm x\{k^2 - C_0^2\}^{\frac{1}{2}}\right] \cosh C_0 y.$$

We are now in a position to describe the condition that we shall impose on ϕ for $x > > 0$. We assume that $\phi(x, y)$ has a wave like form for $x > > 0$. Clearly then, $k^2 < C_0^2$ and we further observe that this condition depends intimately on the values of β and a. As a matter of notation, let us write $\kappa^2 = C_0^2 - k^2$. Then for $x > > 0$, $\phi(x, y)$ is asymptotic to $\exp(\pm i\kappa x) \cosh C_0 y$, while $\Phi_1(x, y, z)$ will be asymptotic to

$$\exp\left[\pm i\kappa x \pm ikz\right] \cosh C_0 y.$$

This asymptotic form of Φ_1 has the character of a plane wave. The normal to its wave front makes an angle with respect to the positive x axis which may be measured in terms of k and κ. Indeed if we write $k = C_0 \sin \alpha$, then κ will be $C_0 \cos \alpha$. The wave front normal will be perpendicular to the z axis, the edge of the dock, if $\alpha = 0$, that is $k = 0$.

In order to obtain a travelling wave solution [13], we find it necessary to admit a logarithmic singularity at the edge of the dock. That is, we shall assume that $\phi(x, y) = 0\{ln(x^2 + (y - a)^2)^{\frac{1}{2}}\}$ in the neighborhood of $x = 0$, $y = a$. This condition is tantamount to assuming that there is either a sink or a source at this point. We shall see in the final solution of our problem that the strength of this sink or source is closely related to the amplitude of the terms $\exp(\pm i\kappa x)$. Incidently, this logarithmic singularity is the strongest singularity which we can admit in our formulation.

Our method of solution is the following. With the aid of an appropriate Green's function and an application of Green's theorem, we may express $\phi(x, y)$ in the strip $0 \leq y \leq a$, $-\infty < x < \infty$ in terms of the Green's function and $\phi(x, a)$. This leads us to an integral equation of the Wiener-Hopf type because of the particular limits of integration and the special x variation of the Green's function. Before we can proceed to the formulation of the integral equation we require an explicit expression for the Green's function. Let us note, in passing, that we have here a mixed boundary value problem. That is, for $x > 0$, $y = a$, we have one type of boundary condition, while on the same line $y = a$ but now $x < 0$, we have a second one. It is for this reason that the Green's function technique plays such an important role here. So to speak, it relieves the surface $y = a$, $x > 0$ or $x < 0$ of boundary conditions.

We shall divide the solution of this problem into two parts. First we shall find the source free solution. Having obtained this solution, it is then possible to obtain the solution possessing a singularity by differentiation of the bounded solution. An appropriate linear combination of these two solutions gives us the traveling wave solution which we desire.

With minor modifications in the formulation which we shall present, we can treat the case of normal incidence, that is $k = 0$. The solution can be carried through in the same fashion as the case for which $k \neq 0$. We shall find, as we might suspect, that we simply put $k = 0$ throughout our final result for $\phi(x, y)$. The condition that $k^2 < C_0^2$, which is required to insure wave motion, is now automatically satisfied.

For the case in which the channel is of infinite depth, we have a somewhat different situation. If $k \neq 0$, we can still solve the problem which we have described above, although the mathematical technique is more subtle than for the case of finite depth. For $k = 0$, our mode of formulation is still valid, but our mathematical machinery breaks down since the Wiener-Hopf theory is no longer applicable. To be more precise, the kernel of the integral equation for $k = 0$ does not possess the exponential decay properties which it does for $k \neq 0$. We shall return to the case of infinite depth at a later date.

2. **The construction of Green's Function** [14]. The nature of the Green's function we require is dictated by the form of equation (1.5) and its associated boundary conditions on $y = 0$ and $y = a$. We are interested in the explicit representation of a function of two sets of coordinates (x, y) and (x', y') which satisfies the partial differential equation

$$G_{xx} + G_{yy} - k^2 G = 0 \tag{2.1}$$

save at the point $x = x'$, $y = y'$. At this point the x and y derivatives of $G(x, y, x', y')$ suffer discontinuities of the form

$$\int_{-\infty}^{\infty} G_y \Big|_{y=y'-0}^{y=y'+0} dx = -1 \quad \text{and} \int_0^a G_x \Big|_{x=x'-0}^{x=x'+0} dy = -1. \tag{2.2}$$

As for boundary conditions we choose the following. We require that

$$G_y = 0 \qquad y = 0, \ -\infty < x < \infty \tag{2.3}$$

and

$$G_y = 0 \qquad y = a, \ -\infty < x < \infty. \tag{2.4}$$

We have a considerable amount of freedom in the choice of these boundary conditions and the ones we have selected lead to a simple form of the integral equation which will be formulated in **3**.

Such a Green's function as we have just described has the expansion

$$\begin{aligned} G(x, y, x', y') &\\ = \sum_{n=0}^{\infty} &\exp[-| x - x' | \{k^2 + (n^2\pi^2/a^2\}^{\frac{1}{2}}] \\ &\times \cos(n\pi y/a) \cos(n\pi y'/a)/\{a^2k^2 + n^2\pi^2\}^{\frac{1}{2}}. \end{aligned} \tag{2.5}$$

The derivation proceeds along lines described by H. Feshbach and the author and we shall not pursue this matter further. Observe that for $x \gg x'$, $G(x, y, x', y')$ is asymptotic to $\exp[|k|(x' - x)]$ while for $x \ll x'$ it is asymptotic to $\exp[|k|(x - x')]$. It is partially because of these asymptotic

properties of $G(x, y, x', y')$ that we can apply Fourier methods to the solution of the integral equation we shall derive.

An important property of the two dimensional Green's function, of the type we have considered, is its logarithmic character near $x = x'$, $y = y'$. In order to see this, we note that the expansion of $G(x, y, x', y')$ converges uniformly and absolutely provided $x \neq x'$. Furthermore for n sufficiently large, the n-th term of the expansion is asymptotic to

$$\{\cos (n\pi y/a) \cos (n\pi y'/a)/n\pi\} \exp [-n\pi \mid x - x' \mid /a].$$

The singular part of $G(x, y, x', y')$ is then determined from the singular part of

$$\sum_{n=1}^{\infty} \{\cos (n\pi y/a) \cos (n\pi y'/a)/n\pi\} \exp [-n\pi \mid x - x' \mid /a].$$

This series may be summed provided that $x \neq x'$ and we get immediately

$$-1/4\pi \log 4 \exp [-2\pi/a \mid x - x' \mid] [\cosh \pi/a (x - x') - \cos \pi/a (y - y')]$$
$$\text{times } [\cosh \pi/a (x - x') - \cos \pi/a \, (y + y')].$$

In the neighborhood of the point $x = x'$, $y = y'$ this has the form

$$-1/4\pi \log \{(x - x')^2 + (y - y')^2\}$$

that is, it possesses a logarithmic singularity.

We close this section with the Fourier integral representation of equation (2.5). Since the expansion for $G(x, y, x', y')$ is asymptotic to $\exp [(x' - x) \mid k \mid]$ for $x >> x'$ and $\exp [(x - x') \mid k \mid]$ for $x << x'$ and since it possesses an integrable singularity at $x = x'$, $y = y'$, we may take the Fourier transform of each term in the expansion and sum them. Upon doing this, we get

$$(2.6) \qquad \int_{-\infty}^{\infty} \exp (-iwx) G(x, y, x', y') \, dx$$
$$= \exp (-iwx') \frac{\cosh \gamma y \cosh \gamma (a - y')}{\gamma \sinh \gamma a}, \qquad y < y'$$
$$= \exp (-iwx') \frac{\cosh \gamma y' \cosh \gamma (a - y)}{\gamma \sinh \gamma a}, \qquad y > y'$$

where $\gamma^2 = k^2 + w^2$ and $-\mid k \mid < \text{Im} w < \mid k \mid$. The expansion (2.5) is obtained immediately by carrying out the Fourier inversion of equation (2.6), that is

$$1/2\pi \int_C \exp [iw(x - x')] \frac{\cosh \gamma y \cosh \gamma (a - y')}{\gamma \sinh \gamma a} \, dw, \qquad y < y'$$

and a similar integral with y and y' interchanged for $y > y'$. The path C is drawn with the strip $-|k| < \text{Im}w < |k|$ of the complex w plane and is closed above or below depending on whether $x > x'$ or $x < x'$. The closing path is a semi-circle of large radius which passes between the poles of the integrand. When the radius of the circle is permitted to become infinite, we can show by methods which are familiar in contour integration, that the semi-circular path makes no contribution to the integral. We are then left with the task of calculating the residues and this gives us (2.5) immediately.

3. Formulation of the Weiner-Hopf integral equation [15], [16]. We can express $\phi(x, y)$ in the strip $-\infty < x < \infty$, $0 < y < a$ in terms of $\phi(x, a)$, its normal derivative and the Green's function which we have just described. Since $\phi(x, y)$ has been assumed to be source free in this strip, we may apply Green's Theorem over a rectangle whose boundaries are $y = 0$, $y = a$, $x = -l$, $x = l_1$ $(l, l_1 >> 0)$ and we get

$$\phi(x, y) = \int [G(x, y, x', y')(\partial\phi(x', y')/\partial n') \\ - \phi(x', y')(\partial G(x, y, x', y')/\partial n')]ds'$$

where ds' is the element of arc-length along the boundary of the rectangle which we have described. The operation $\partial/\partial n'$ denotes the outer normal derivative. It is to be noted that this integral is treated as an improper one in the sense that we may allow l and l_1 to become infinitive after we have discussed the magnitude of the integrals along these boundaries. We are permitted to do this because of the decay properties of $G(x, y, x', y')$ for $x >> x'$ or $x << x'$ and the assumption regarding $\phi(x, a)$ for $|x| >> 0$.

Over the lower boundary $y = 0$, there are no contributions in view of the boundary conditions imposed on ϕ_y and G_y. On the upper boundary we have

$$\int_0^{l_1} G(x, y, x', a)\phi_{y'}(x', a)\,dx'. \tag{3.1}$$

Upon noting the boundary conditions on $\phi(x, y)$ for $y = 0$, $x > 0$, and the nature of its growth for $x >> 0$, we may let l_1 become infinite and (3.1) becomes

$$\beta \int_0^{\infty} G(x, y, x', a)\phi(x', a)\,dx'.$$

Let us now examine the integrals over the lines $-l$ and l_1. Over the left boundary we have that $x >> x'$. Then $G(x, y, x', y')$ is asymptotic to

$\exp [(x'-x)|k|]$. Further $\phi(x,y)$ is also asymptotic to $\exp [|k|x]$ for $x \ll 0$. Hence

$$\int_0^a [G(x,y,x',y')\phi_{x'}(x',y') - \phi(x',y')G_{x'}(x,y,x',y')]\,dy'$$

approaches zero uniformly when l becomes infinite. A similar remark may be addressed to the integral on the right. Hence, for any (x,y) in the infinite strip we have

$$\phi(x,y) = \beta \int_0^\infty G(x,y,x',a)\phi(x',a)\,dx' \tag{3.2}$$

and upon evaluating $\phi(x,y)$ at $y=a$, we get the desired Wiener-Hopf integral equation, that is

$$\phi(x,a) = \beta \int_0^\infty G(x,a,x',a)\phi(x',a)\,dx'. \tag{3.3}$$

The integral equation (3.3) now includes all of the boundary conditions at infinity. This equation, as we shall see, possesses a solution of the form which acts like $\eta_1 + \eta_2 x \log x$ for $x \to 0^+$, (η_1 and η_2 are independents of x) and is bounded at infinity. From this, by a differentiation with respect to x, we can find the solution which acts like $\log x$, or indeed by further differentiation, solutions which act like x^{-n}, $n = 1, 2, \cdots x \to 0^+$ [17]. We now proceed to the solution of equation (3.3).

4. The Fourier transform solution of equation (3.3). Following Wiener and Hopf, we define $\phi(x,a) = \phi_1(x) + \phi_2(x)$ where $\phi_1(x)$ is identically zero for negative x and $\phi_2(x)$ is identically zero for positive x. Equation (3.3) may be rewritten as

$$\phi_1(x) + \phi_2(x) = \beta \int_{-\infty}^\infty G(x,a,x',a)\phi_1(x')\,dx'. \tag{4.1}$$

From the functional form of $G(x,a,x',a)$ it is evident that $\phi_2(x)$ is asymptotic to $\exp [|k|x]$ for x large and negative. In order to apply Fourier techniques to the solution of equation (4.1) we are required to know the regions of regularity of the Fourier transforms of $G(x,a,x',a)$, $\phi_1(x)$ and $\phi_2(x)$. In the first place, the Fourier transform of $G(x,a,x',a)$ is

$$\int_{-\infty}^\infty G(x,a,x',a)\exp(-iwx)\,dx = \exp(-iwx')\coth \gamma a/\gamma$$

and as we have defined it,the transform is regular in the strip $-|k| < \mathrm{Im} w < |k|$ in the complex w plane. As for the transform of $\phi_2(x)$ we have

$$\int_{-\infty}^{0} \phi_2(x) \exp(-iwx)\, dx = \Phi_2(w)$$

is regular in the upper half plane $\mathrm{Im} w > -|k|$. This is so because of the growth of $\phi_2(x)$ for x large and negative. It has been assumed that $\phi_2(x)$ is integrable over any finite interval of the negative x axis including the origin, and this we shall verify in the solution of the problem. As for $\phi_1(x)$, we shall assume that it is $O[\exp(\theta x)]$, $\theta < |k|$. We shall find that $\phi_1(x)$ does not grow as rapidly as this, but this is all that we require to solve equation (4. 1) by Fourier methods. The Fourier transform of $\phi_1(x)$,

$$\int_{0}^{\infty} \phi_1(x) \exp(-iwx)\, dx = \Phi_1(w)$$

is now regular in the lower half plane $\mathrm{Im} w < -\theta$. We note that the above three Fourier transforms have a common region of regularity $-|k| < \mathrm{Im} w < -\theta$ and it is thus permissible to apply the Fourier transform to equation (4. 1).

Upon doing this we obtain immediately

(4. 2) $$\Phi_1(w) + \Phi_2(w) = \Phi_1(w)\, \beta \coth \gamma a/\gamma$$

or

(4. 2a) $$\Phi_1(w)[1 - \beta \coth \gamma a/\gamma] = -\Phi_2(w).$$

We are now required to decompose equation (4. 2a) into two terms, one analytic in the lower half plane $\mathrm{Im} w < -\theta$, the other analytic in the upper half plane $\mathrm{Im} w > -|k|$, while both terms are analytic in a common strip $-|k| < \mathrm{Im} w < -\theta$. To accomplish this, we factor

(4. 3) $$(\gamma \sinh \gamma a - \beta \cosh \gamma a)/\gamma \sin \gamma a = K_-(w)/K_+(w).$$

$K_-(w)$ is that factor of the left side of equation (4. 3) which is free of zeros and poles in the lower half plane $\mathrm{Im} w < 0$. $K_+(w)$ may be defined in terms of the left side of equation (4. 3) and the factor $K_-(w)$. It is free of zeros and poles in the upper half plane $\mathrm{Im} w > -|k|$. Hence

(4. 4) $$\Phi_1(w) K_-(w) = -\Phi_2(w) K_+(w).$$

The left side of equation (4. 4) is analytic in the lower half plane $\mathrm{Im} w < 0$ while the right side is analytic in the upper half plane $\mathrm{Im} w > -|k|$. Hence the left side is the analytic continuation of the right side in view of the common strip of analyticity and hence both sides are regular everywhere. That is, each side of equation (4. 4) is equal to a common integral function $E(w)$. We have then

(4. 5a) $$\Phi_1(w)K_-(w) = E(w)$$
and
(4. 5b) $$\Phi_2(w)K_+(w) = -E(w).$$

The determination of the integral function $E(w)$ depends on the asymptotic forms of $\Phi_1(w)$, $\Phi_2(w)$, $K_-(w)$ and $K_+(w)$ so that we find it necessary to exhibit $K_-(w)$ and $K_+(w)$ explicitly.

We have noted that $K_-(w)$ is free of zeros and poles in the lower half plane $\mathrm{Im} w < 0$. The zeros of $\gamma \sinh \gamma a - \beta \cosh \gamma a = 0$ as a function of γa are $\pm \rho_0$ and $\pm i\rho_n$, $\rho_n = -\rho_{-n}$: $n = 1, 2, \cdots$ where the ρ_n are real. We observe in passing that for $n >> 0$,

$$\rho_n = n\pi + \beta a/n\pi.$$

In order to simplify the writing of $K_-(w)$ and $K_+(w)$ we write

$$K(w) = L(w)/M(w)$$

where

$$L(w) = \gamma \sinh \gamma a - \beta \cosh \gamma a$$

and

$$M(w) = \gamma \sinh \gamma a.$$

Now $L(w)$ may be written in factor form as

$$-\beta[1 - \gamma^2 a^2/\rho_0{}^2] \prod_{n=1}^{\infty} [1 + \gamma^2 a^2/\rho_n{}^2]$$
$$= -\beta[1 - a^2(k^2 + w^2)/\rho_0{}^2] \prod_{n=1}^{\infty} [1 + a^2(k^2 + w^2)/\rho_n{}^2].$$

Hence

$$L_-(w) = [1 - a^2(k^2 + w^2)/\rho_0{}^2] \prod_{n=1}^{\infty} [\{1 + a^2k^2/\rho_n{}^2\}^{1/2} + iaw/\rho_n]\exp[-iaw/n\pi]$$

is free of zeros in the lower half plane $\mathrm{Im} w < 0$, while

$$1/L_+(w) = -\beta \prod_{n=1}^{\infty} [\{1 + a^2k^2/\rho_n{}^2\}^{1/2} - iaw/\rho_n]\exp[iaw/n\pi]$$

if free of zeros in the upper half plane

$$\mathrm{Im} w > -\{\rho_1/a\}\{1 + a^2k^2/\rho_1{}^2\}^{1/2}.$$

The exponential factors have been inserted into the infinite products to insure their absolute convergence. Finally,

$$L(w) = L_-(w)/L_+(w).$$

Similarly for the function $M(w)$, we have

$$M_-(w) = a(k + iw) \prod_{n=1}^{\infty} [\{1 + a^2k^2/n^2\pi^2\}^{\frac{1}{2}} + iaw/n\pi]\exp[-iaw/n\pi]$$

is free of zeros in the lower half plane $\text{Im}w < |k|$, while

$$1/M_+(w) = M_-(-w)/a$$

is free of zeros in the upper half plane $\text{Im}w > -|k|$. Hence

$$K_-(w) = L_-(w)/M_-(w)$$

is regular in the lower half plane $\text{Im}w < 0$. We imply by regularity in this case that $K_-(w)$ and its reciprocal are regular in the above described half plane. Finally,

$$1/K_+(w) = M_+(w)/L_+(w)$$

is regular in the upper half plane $\text{Im}w > -|k|$.

We are now in a position to determine the integral function $E(w)$. In the first place $K_-(w) = O(w)$ for $\text{Im}w < 0$, $|w| \to \infty$ in view of the asymptotic nature of the ρ_n's. Similarly $K_+(w) = O(w)$ for $\text{Im}w > -|k|$, $|w| \to \infty$. Now $\Phi_1(w)$ is analytic in the lower half plane $\text{Im}w < 0$ and approaches zero for $|w| \to \infty$ in this half plane. Thus, in the lower half plane $E(w) = O(w^{\sigma_1})$, $\sigma_1 < 1$. Furthermore $E(w) = O(w^{\sigma_2})$, $\sigma_2 < 1$ in the appropriate upper half plane. Since $E(w)$ is an integral function, σ_1 and σ_2 are integers and necessarily greater than or equal to zero. In this case they are both zero because of the inequality they both satisfy. Hence $E(w)$ is a constant B. We have then

(4. 6a) $$\Phi_1(w) = B/K_-(w)$$

(4. 6b) $$\Phi_2(w) = -B/K_+(w).$$

5. The calculation of $\phi(x, y)$. In order to calculate $\phi(x, y)$, we write the integral equation (3. 2) in Fourier integral form. We recall that $\phi(x, y)$ is that solution of our boundary value problem which is bounded at the point $x = 0$, $y = a$. Thus

(5. 1) $$\phi(x, y) = \beta/2\pi \int_C \frac{\exp(iwx) \cosh \gamma y\, \Phi_1(w)\, dw}{\gamma \sinh \gamma a}$$

where the path C is drawn in the strip $-|k| < \text{Im}w < 0$. It is closed above or below depending on whether $x > 0$ or $x < 0$. Upon substituting (4. 6a) into (5. 1) we have

(5.2a) $$\phi(x, y) = \beta B/2\pi \int_C \frac{\exp(iwx)\cosh\gamma y}{\gamma\sinh\gamma a}\frac{dw}{K_-(w)}$$

(5.2b) $$= \beta B/2\pi \int_C \frac{\exp(iwx)\cosh\gamma y}{(\gamma\sinh\gamma a - \beta\cosh\gamma a)}\frac{dw}{K_+(w)}$$

The representation (5.2a) is particularly appropriate for $x < 0$, since in that case C is closed below and it is now a simple task to calculate the residues at the poles $w = -i\{k^2 + n^2\pi^2/a^2\}^{1/2}$, $n = 0, 1, \cdots$. Similarly, (5.2b) is appropriate for $x > 0$. For $x > 0$, we have

(5.3) $$\phi(x, y) = B\sum_{n=1}^{\infty} \frac{\beta\rho_n^2\{\cos\rho_n y/a\}\exp[-x\{\rho_n^2/a^2 + k^2\}^{1/2}]}{\{\rho_n^2 + a^2k^2\}^{1/2}(\rho_n^2 + \beta^2a^2 - \beta a)\cos\rho_n K_+[i\{k^2 + \rho_n^2/a^2\}^{1/2}]} + \frac{2^{1/2}\beta B\rho_0^2\{\cosh\rho_0 y/a\}\sin(\sigma x + \Theta)}{\sigma\cosh\rho_0\{a\beta - a^2\beta^2 + \rho_0^2\}^{1/2}}.$$

The angle Θ is given by the infinite sums

$$\Theta = -\sum_{n=1}^{\infty}[\arcsin a\sigma(\rho_n^2 + \rho_0^2)^{-1/2} - a\sigma/n\pi] + \sum_{n=1}^{\infty}[\arcsin a\sigma(n^2\pi^2 + \rho_0^2)^{-1/2} - a\sigma/n\pi] + \arcsin a\sigma/\rho_0.$$

σ is simply the positive root of the equation

$$a^2k^2 + a^2\sigma^2 = \rho_0^2$$

and is real under our assumption that ρ_0 exceeds ak. For $x < 0$, we have

(5.4) $$\phi(x, y) = \beta B\left[\frac{\exp|k|x}{2a|k|K_-(-i|k|)} + \sum_{n=1}^{\infty}\frac{\exp[\{k^2 + n^2\pi^2/a^2\}^{1/2}x](-)^n\cos n\pi y/a}{(a^2k^2 + n^2\pi^2)^{1/2}K_-[-i(k^2 + n^2\pi^2/a^2)^{1/2}]}\right].$$

These two forms may be obtained immediately by the calculation of the residues in the integral in equation (5.1) by appropriate closing of C and noting that there is no contribution from the closing path which we have already described in **2**.

Since

$$K_-[-i\{n^2\pi^2/a^2 + k^2\}^{1/2}] \text{ and } K_+[i\{\rho_n^2/a^2 + k^2\}^{1/2}]$$

are asymptotic to $n\pi$, the infinite series contributions in equations (5.3) and (5.4) have coefficients which are $O(n^{-2})$ for n sufficiently large. The series will then converge uniformly and absolutely for $x \geqq 0$ or $x \leqq 0$,

depending upon which representation of $\phi(x, y)$ we employ. Thus $\phi(x, y)$ is integrable near $x = 0^-$ and it is clearly integrable for x large and negative. Similar remarks may be addressed to the function $\phi(x, y)$ which is defined for $x > 0$. For a fixed x, expansions (5.3) and (5.4) represent the developments of $\phi(x, y)$ in the appropriate characteristic functions of the regions $0 \leq y \leq a$, $x \leq 0$ and $0 \leq y \leq a$, $x \geq 0$. The expansion coefficients described in **1** have been determined only up to a constant of proportionality B because of the homogeneous character of equation (3.3).

6. The solution with the logarithmic singularity. From equation (5.3) and (5.4) it is a direct step to the solution which possesses the logarithmic singularity. Since the series expansions converge uniformly and absolutely for all y in $0 \leq y \leq a$ and for $x > 0$ or $x < 0$ as the case may be, we may differentiate either of them term by term with respect to x. For $x > 0$ we have

$$\phi_x = -\sum_{n=1}^{\infty} \frac{B\beta\rho_n^2\{\cos \rho_n y/a\} \exp\left[-x\{\rho_n^2/a^2 + k^2\}^{1/2}\right]}{(a\cos\rho_n)(\rho_n^2 + \beta^2a^2 - \beta a)K_+[i\{\rho_n^2/a^2 + k^2\}^{1/2}]} + \frac{\sqrt{2}\,B\beta\rho_0^2\{\cosh \rho_0 y/a\}\cos(\sigma x + \Theta)}{(\cosh\rho_0)\{a\beta - a^2\beta^2 + \rho_0^2\}^{1/2}}.$$

ϕ_x is still a solution of equation (4.5) with the same type of growth for $x \gg 0$ as $\phi(x, y)$. It satisfies the boundary conditions on the boundaries $y = a$ and $y = a$ because the expansion still converges uniformly and absolutely for $0 \leq y \leq \mathrm{a}$, $x > 0$. For $x \to 0^+$, $y = a$, ϕ_x possesses a logarithmic singularity. Hence in order to obtain the solution of equation (1.5) which possesses a bounded as well as a logarithmic component, it is merely necessary to take a linear combination of $\phi(x, y)$ and ϕ_x. With appropriate constants we can settle the strength of the logarithmic source. We shall discuss the properties of ϕ and ϕ_x in more detail in **7**.

7. Some properties of $\phi(\mathrm{x}, \mathrm{y})$, $\phi_1(\mathrm{x})$ and $\phi_2(\mathrm{x})$. We can show from the integral representation of $\phi(x, y)$ that $\phi(0^+, y) = \phi(0^-, y)$ for $0 \leq y \leq a$ and hence that $\phi_1(0) - \phi_2(0)$ since $\phi_1(x) = \phi(x, a)$, $x > 0$ and $\phi_2(x) = \phi(x, a)$ $x < 0$. Furthermore we shall show that

$$\phi_x \Big|_{x=0^+} = \phi_x \Big|_{x=0^-} \qquad 0 \leq y < a.$$

The point $x = 0$, $y = a$ requires special discussion and indeed we know that ϕ_x has a logarithmic singularity at this point. We recall that

$$(7.1) \qquad \phi(x, y) = \beta B/2\pi \int_C \frac{\exp(iwx) \cosh \gamma y \, dw}{\gamma \sinh \gamma a \, K_-(w)} \qquad \gamma^2 = k^2 + w^2.$$

The path C has been drawn in the strip $-|k| < \operatorname{Im} w < 0$. Now $K_-(w) = O(w)$, $|w| \to \infty$ if w is exterior to a sector which includes the positive imaginary axis of the w plane. That is, since $\rho_n = n\pi + \beta a/n\pi$ for n sufficiently large, the ratio of the infinite products in $K_-(w)$ will be constant for $|w| \to \infty$, provided w is exterior to the sector which includes the imaginary zero and poles in the upper half plane. It is then possible to deform the path C so that the path starts from the left at infinity on a straight line making a slope $\pi - \lambda$ with the positive u axis $(w = u + iv)$ passes to the left of $w = -\sigma$, thence below the u axis into the strip $-|k| < \operatorname{Im} w < 0$, then to the right of $w = \sigma$ and finally to the line which makes a slope λ with positive u axis and to infinity $(0 < \lambda < \pi/2)$. This deformation is appropriate for $x \geq 0$ since we now can write $w = \rho \exp i\lambda$ and equation (7.1) is of the order [18]

$$\int^{\infty} \exp [\rho(y - a) \cos \lambda - \rho x \sin \lambda] d\rho/\rho^2.$$

This integral converges uniformly for all $x \geq 0$ and $0 \leq y \leq a$. Hence equation (7.1) defines a continuous function of x for $x \geq 0$ and $0 \leq y \leq a$. Similarly ϕ_x is of the order

$$\int^{\infty} \exp [\rho(y - a) \cos \lambda - \rho x \sin \lambda] d\rho/\rho$$

for $x > 0$. Now we see that the range of validity is modified. If $x > 0$, the integral converges uniformly in the interval $0 \leq y \leq a$, and if $0 \leq y < a$, for all $x \geq 0$. Hence ϕ_x is a continuous function of x if $0 \leq y < a$, and $x \geq 0$ or $0 \leq y \leq a$ and $x > 0$. In fact we note that any derivative of ϕ is continuous as long as $0 \leq y \leq a$, $x > 0$ or $0 \leq y < a$, $x \geq 0$.

For $x < 0$, the appropriate deformation is in the lower half plane $\operatorname{Im} w < 0$. The sides of the sector now have the slope $\pi + \lambda$ and $2\pi - \lambda$. We simply reverse all x inequalities in the above paragraph and we have the desired result. We see therefore that $\phi(x, y)$ is continuous for $-\infty < x < \infty$, $0 \leq y \leq a$ and hence $\phi(0_+, y) = \phi(0_-, y)$. Furthermore, any partial derivative of $\phi(x, y)$ is continuous provided either (i) $0 \leq y \leq a$, $x \gtrless 0$ or (ii) $0 \leq y < a$, $x \geq 0$ or $x \leq 0$. Hence, since the integral representation has been used to deduce the series representations for $\phi(x, y)$ in a rigorous fashion, we see that the series may be differentiated twice with respect to x or y and be shown to satisfy the differential equation (1.5), for $0 \leq y < a$,

$-\infty < x < \infty$, or $0 \leq y \leq a$, $x \gtrless 0$. Similarly the boundary conditions may be satisfied. We have thus verified that the $\phi(x,y)$ we have derived is indeed bounded in the strip and satisfies the differential equation (1.5) except in the neighborhood of the point $x = 0$, $y = a$.

We shall now discuss the behavior of ϕ_x and ϕ_y in the neighborhood of $x = 0$, $y = a$. In the first place

$$1/K_-(w) = [1/w][\alpha_1 + (\alpha_2 \log w)/w]$$

for $|w| \to \infty$, Im$w < 0$, where α_1 and α_2 are independent of w [11]. On the other hand

$$1/K_+(w) = -a\beta\rho_0^2 K_-(-w)/[\rho_0^2 - a^2(k^2 + w^2)]$$

so that for $|w| \to \infty$, Im$w > -|k|$

$$1/K_+(w) = -[\rho_0^2\beta/a\alpha_1 w][1 + (\alpha_2/\alpha_1)(\log - w)/w].$$

The constant α_1 may be determined by noting that

$$(\gamma \sinh \gamma a - \beta \cosh \gamma a)/\gamma \sinh \gamma a = K_-(w)/K_+(w) = K(w)$$

has the limit unity as $|w| \to \infty$ in the strip $-|k| < \text{Im}w < 0$. We have then

$$K_-(w)/K_+(w) = -[\beta\rho_0^2/a\alpha_1^2][1 + O(\log w/w)]$$

in this strip and hence

$$-\beta\rho_0^2 = a\alpha_1^2.$$

This implies that α_1 is imaginary. We observe furthermore that

$$K_-[-i\{k^2 + n^2\pi^2/a^2\}^{1/2}] = n\pi/i\alpha_1 a, \qquad n \gg 0$$

and

$$K_+[i\{k^2 + \rho_n^2/a^2\}^{1/2}] = n\pi\alpha_1/i\rho_0^2\beta = -n\pi/i\alpha_1 a, \qquad n \gg 0.$$

The behavior of $\phi(x,y)$ in the neighborhood of $x = 0^+$, $y = a$ may be derived from the series (5.3) in the following fashion. For n sufficiently large, the n-th term of (5.3) is asymptotic to

$$-[i\alpha_1 a^2 B\beta(-)^n/(n\pi)^2][\cos n\pi y/a] \exp[-n\pi x/a]$$

and the sum of such terms from $n = 1$ to ∞ converges uniformly and absolutely for $x \geq 0$. In order to indicate the behavior of this series in the neighborhood of the point $x = 0^+$, $y = a$, we write

$$\zeta = \pi[x + i(a - y)]/a \text{ and } \bar{\zeta} = \pi[x - i(a - y)]/a.$$

Then for $x = 0^+$, $y = a$

$$\begin{aligned}\phi(x,y) &= -(i\alpha_1 a\beta B/2\pi^2)\sum_{n=1}^{\infty}[\exp(-n\zeta)+\exp(-n\bar{\zeta})]/n^2 + C' \\ &= -(i\alpha_1 a\beta B/2\pi^2)[\zeta\log\zeta+\bar{\zeta}\log\bar{\zeta}] + C\end{aligned}$$

where C and C' are constants [17]. Similarly for $x = 0^-$, $y = a$, we get that

$$\begin{aligned}\phi(x,y) &= (i\beta B\alpha_1 a/\pi^2)\sum_{n=1}^{\infty}\exp(n\pi x/a)(-)^n(\cos n\pi y/a)/n^2 + C_1' \\ &= (i\beta B\alpha_1 a/2\pi^2)\sum_{n=1}^{\infty}[\exp(n\zeta)+\exp(n\bar{\zeta})]/n^2 + C_1' \\ &= (i\beta B\alpha_1 a/2\pi^2)(-\zeta\log-\zeta-\bar{\zeta}\log-\bar{\zeta}) + C_1\end{aligned}$$

where C_1 and C'_1 are constants. The continuity requirement at $x = 0$, $y = a$ on the function ϕ gives us $C_1 = C$. The x derivative of $\phi(x,y)$ in the neighborhood of $x = 0$, $y = a$ exhibits a singularity. That is, in the neighborhood of this point

$$\phi_x = (-i\alpha_1 a\beta B/2\pi)\log\zeta\bar{\zeta}$$

and hence there is a logarithmic singularity there. This tells us, that while $\phi(x,y)$ is everywhere bounded in the strip $-\infty < x < \infty$, $0 \leq y \leq a$, its x derivative has a singularity, in this case a two dimensional source or sink. We also note that for $x \gg 0$, $\phi(x,y)$ and ϕ_x are out of phase by an angle of $\pi/2$ radians. Upon reinserting the time factor which has been omitted and forming an appropriate linear combination of these solutions, we can find a travelling wave solution for this problem [10]. The amplitude for $x \gg 0$ depends on the strength of the source as we shall see in the next section.

8. A reciprocity theorem. We have remarked in **1** that the strength of the sink or source in ϕ_x is related to the amplitude of the dominant term in $\phi(x,y)$ as $x \to \infty$. In order to find this relation, we write $\psi(x,y) = \phi_x$ where $\phi(x,y)$ is the bounded solution and therefore $\psi(x,y)$ has a logarithmic singularity, provided $\alpha_1 \neq 0$. From Green's theorem we have

$$(8.1)\qquad \begin{aligned}&\int[\phi(x',y')\partial\psi(x',y')/\partial n' - \psi(x',y')\partial\phi(x',y')\partial n']dS' \\ &= \int\int[\phi\nabla^2\psi - \psi\nabla^2\phi]dA\end{aligned}$$

where the integrals are taken over a rectangle whose boundaries are $x = -l$, $x = l_1 (l, l_1 \gg 0)$ $y = 0$, $y = a$. Since ψ and ϕ_n have logarithmic singularities, we are compelled to make a small circular indentation at $x = 0$, $y = a$ and

take the limit as the radius of the indentation goes to zero. The surface integral in equation (8.1) vanishes since $\nabla^2\phi = k^2\phi$ and $\nabla^2\psi = k^2\psi$. Hence we simply need calculate the line integral

$$\text{(8.1a)} \qquad \int (\phi\partial\psi/\partial n' - \psi\partial\phi/\partial n')\, dS' = 0.$$

Now ϕ_n and ψ_n vanish at $y = 0$, $-l < x < l_1$ and $y = a$, $-l < x < 0$. Furthermore for $y = a$, $x > 0$, $\phi_n = \beta\phi$ and $\psi_n = \beta\psi$. Equation (8.1a) then reduces to an integral over the indentation and the lines $x = -l$ and $x = l_1$. But for l sufficiently large, ϕ, ϕ_x, ψ, and ψ_x are $O[\exp(-\pi l/a)]$ so that for $l \to \infty$, there is no contribution from the left side of the rectangle. For l_1 sufficiently large

$$\phi(l_1, y) \doteq \frac{2^{1/2}\beta B\rho_0{}^2 \cosh(\rho_0 y/a)\sin(\sigma l_1 + \Theta)}{\sigma \cosh\rho_0\{a\beta - a^2\beta^2 + \rho_0{}^2\}^{1/2}} + O[\exp(-\pi l_1/a]$$

and

$$\psi(l_1, y) = \frac{2^{1/2}\beta B\rho_0{}^2 \cosh(\rho_0 y/a)\cos(\sigma l_1 + \Theta)}{\cosh\rho_0\{a\beta - a^2\beta^2 + \rho_0{}^2\}^{1/2}} + O[\exp(-\pi l_1/a)].$$

The contribution at $x = l_1$ is

$$\frac{-2\beta^2 B^2 \rho_0{}^4}{\cosh^2\rho_0(a\beta - a^2\beta^2 + \rho_0{}^2)}\int_0^a \cosh^2(\rho_0 y/a)\, dy + O[\exp(-\pi l_1/a)]$$

or

$$-a\beta^2 B^2 \rho_0{}^2 + O[\exp(-\pi l_1/a)].$$

When l_1 becomes infinite we simply have $-a\beta^2 B^2\rho_0{}^2$.

We are now left with the task of computing the contribution from the indentation. As a matter of notation we write $r^2 = x^2 + (a - y)^2$, $x = r\cos\theta$, and $a - y = r\sin\theta$. Then we have on the indentation

$$-\int_0^\pi (Ci\alpha_1 a\beta B/r\pi) r\, d\theta + O(r\log r) = -Ci\alpha_1 a\beta B$$

when $r \to 0$. We have finally

$$\text{(8.2)} \qquad Ci\alpha\beta B = a\beta^2 B^2 \rho_0{}^2$$

or

$$Ci\alpha_1 = \beta B\rho_0{}^2,$$

a relation between the strength of the source and the amplitudes of the dominant terms in $\phi(x, y)$ and $\psi(x, y)$.

9. Conclusion. In the formulation and the solution of this problem we have used what we may call an "exponential technique." That is, the integral equation was formed with the aid of a Green's function. This Green's function possesses exponential decay properties for $x >> x'$ or $x << x'$. This formulation requires that we make some assumptions regarding the growth of $\phi_1(x)$ and $\phi_2(x)$ for $|x| >> 0$. We found that if we assumed that $\phi_1(x)$ grew no more rapidly than $(\exp \theta x, \theta < |k|)$, and $\phi_2(x)$ no more rapidly than $\exp(|k| x)$ for $x > 0$ or < 0 as the case might be, it was indeed possible to form the integral equation. The $\phi(x, y)$ which we obtained falls within this growth order, that is $\phi_1(x) = O(\exp(\pm i\sigma x))$ for $x >> 0$, while $\phi_2(x) = O(\exp(k | x))$, $x << 0$. On the other hand, we might say that if we stay sufficiently far away from the edge of the dock $(x = 0)$ we can study $\phi(x, y)$ for $|x| >> 0$ by studying two individual problems in potential theory. The first case is that of an infinite dock over a channel of finite depth while the second one is the case of an infinite free surface over a channel of finite depth. These separate problems have established solutions and the growths of $\phi_1(x)$ and $\phi_2(x)$ for $|x| \rightarrow \infty$ would agree with that which we obtained from the integral equation.

A second aspect of the exponential technique is the Fourier transform in the complex domain. Here, we employ a transformation which can overcome the exponential growths of the various functions of which we have to take transforms, provided of course $\operatorname{Im} w$ is suitably limited. We found a $\phi(x, y)$ which did indeed possess these exponential properties and we are now confronted with the question of uniqueness. The general uniqueness theorem will not be discussed here. We simply mention that as a consequence of the general Wiener-Hopf theory, there are no other solutions of this exponential class other than those which we have found. Since we are interested only in such solutions we shall not pursue this matter further.

It is clear that if we are to put our results to use, we should be able to calculate numerically the functions $\phi_1(x)$ and $\phi_2(x)$. This in turn requires that we be able to calculate $K_+[i\{k^2 + \rho_n{}^2/a^2\}^{1/2}]$ and $K_-[-i\{k^2 + n^2\pi^2/a^2\}^{1/2}]$. This may be done without any great difficulty by appealing to the infinite product representations of $K_+(w)$ and $K_-(w)$ and the asymptotic form of ρ_n for n sufficiently large. Complete numerical data and the calculational technique will be published elsewhere at a later date.

CARNEGIE INSTITUTE OF TECHNOLOGY.

BIBLIOGRAPHY AND NOTES.

1. Presented to the American Mathematical Society, April 26, 1947.

2. A. Weinstein, "Sur un probleme mixte pour les fonctions harmoniques," *Rendiconti d'Accademia dei Lincei*, vol. V (ser. 6ᵃ) (1927), pp. 259-265, and "Sur un probleme aux limites dans une indéfinie," *Comptes Rendus* (Paris) vol. 184 (1927), pp. 497-499.

3. G. Hoheisel, "Randwertaufgaben und funktionale Differentialgleichungen," *Jahresbericht der Deutschen Mathematiker Vereinigung*, vol. 39 (1930), pp. 54-58.

4. S. Bochner, *Vorlesungen über Fouriersche Integrale*, Leipzig, 1932, p. 167.

5. H. Poritsky, "On a mixed boundary condition for harmonic functions," *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 723-725.

6. J. Cooper, "A mixed boundary value problem," *Journal of the London Mathematical Society*, vol. 14 (1939), pp. 124-128.

7. A. Heins, "A mixed boundary value problem; some remarks on a problem of A. Weinstein," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 130-133.

8. M. Miche, "Mouvements ondulatoires de la mer en profondeur constante ou décroissante," *Annales des pontes et chaussées*, vol. 114 (1944), pp. 25-78, pp. 131-164, pp. 270-292 and pp. 369-406.

9. H. Lewy, "Water waves on sloping beaches," *Bulletin of the American Mathematical Society*, vol. 42 (1946), pp. 737-775.

10. J. Stoker, "Surface waves in water of variable depth," *Quarterly of Applied Mathematics*, vol. 5 (1947), pp. 1-54.

11. H. Feshbach and A. Heins, unpublished.

12. It is assumed here that the dock has no thickness. The effect of finite thickness of the dock complicates the formulation of the problem, and our method of solution is no longer directly applicable.

13. J. Stoker, *loc. cit.*

14. For further details see A. Sommerfeld, "Die Greensche Funktion der Schwingungsgleichung," *Deutsche Mathematiker Vereinigung, Jahresbericht*, vol. 21 (1912), pp. 309-353.

15. For a discussion of the Wiener-Hopf integral equation see Paley and Weiner, *The Fourier Transform in the Complex Domain*, Colloquium Lectures, American Mathematical Society 1934, Chapter IV.

16. Applications of this integral equation have been given by J. S. Schwinger, "Theory of Guided Waves," Radiation Laboratory Publication, forthcoming; E. T. Copson, "On an integral equation arising in the theory of diffraction," *Oxford Quarterly Journal of Mathematics*, vol. 17 (1946), pp. 19-34; J. F. Carlson and A. E. Heins, "The reflection of an electromagnetic plane wave by an infinite set of plates I," *Quarterly of Applied Mathematics*, vol. 4 (1947), pp. 313-329; A. E. Heins and J. F. Carlson, "The reflection of an electromagnetic plane wave by an infinite set of plates II," *Quarterly of Applied Mathematics*, vol. 5 (1947), p. 82; H. Feshbach and A. E. Heins, *loc. cit.*

17. Differentiation with respect to x does not affect the boundary conditions at $y=0$ or $y=a$. The point $(0,a)$ is to be avoided, of course.

18. Similar deformations are fully described by Carslaw and Jaeger, "Some two-dimensional problems in conduction of heat with circular symmetry," *Proceedings of the London Mathematical Society*, 2nd Ser., vol. 40 (1939-40), p. 361.

19. E. C. Titchmarsh, *Theory of Functions*, Chapter 7, p. 242.

DIRECT DETERMINATION OF BENDING AND TWISTING MOMENTS IN THIN ELASTIC PLATES.*

By H. J. Greenberg and W. Prager.

1. Introduction. A new method of obtaining approximate solutions for a large class of boundary-value problems in Elasticity has recently been developed by W. Prager and J. L. Synge.[1] These authors regard a state of stress in a body of given shape as a point in function space and use the concept of strain energy to establish a metric. The *natural* state of stress which constitutes the solution of the boundary-value problem under consideration satisfies certain boundary conditions, the equations of equilibrium, and the equations of compatibility. Approximations to this natural state of stress are constructed from *artificial* states of stress which satisfy only some of these equations.

The present paper extends this analysis to the bending of thin elastic plates and gives a method for the *direct* determination of bending and twisting moments. From the point of view of the mathematical theory of elasticity, the technical theory of the bending of thin plates is of approximate character. Accordingly, the results of Prager and Synge, which are based on the mathematical theory of elasticity, can not be directly applied to the determination of the moments in a bent plate. To derive analogous results applicable to the theory of plates one must follow the reasoning of Prager and Synge *ab initio* making the necessary modifications at each step.

2. Basic equations. It is convenient to choose the middle surface of the undeformed plate as the x_1, x_2 plane of a system of rectangular Cartesian coordinates x_1, x_2, x_3. In the following, the usual summation convention regarding repeated subscripts is adopted, with Latin subscripts indicating the range 1, 2, 3 and Greek subscripts the range 1, 2.

The plate is subjected to a distributed load $p(x_1, x_2)$ which acts in the direction of the x_3 axis. As is customary in the technical theory of plates,

* Received September 4, 1947. This paper is based on two reports prepared under a contract in Applied Mechanics for Watertown Arsenal. The first of these reports, by H. J. G. and W. P., corresponds to Secs. 1 to 5 of the present paper, the second, by H. J. G., to Sec. 6.

[1] W. Prager and J. L. Synge, "Elasticity and function space," *Quarterly of Applied Mathematics*, vol. 5 (1947), pp. 241-269.

the stress components σ_{i3} are disregarded, and the stress components $\sigma_{\alpha\beta}$ are assumed to be proportional to x_3. The moments $M_{\alpha\beta}$ are then defined by

$$M_{\alpha\beta} = \int_{-h/2}^{h/2} x_3 \sigma_{\alpha\beta} dx_3 \tag{2. 1}$$

where h denotes the thickness of the plate.[2]

With these notations the *equation of equilibrium* takes the form

$$M_{\alpha\beta,\alpha\beta} + p = 0, \tag{2. 2}$$

where a subscript appearing after a comma indicates differentiation with respect to the corresponding coordinate.

If $w = w(x_1, x_2)$ denotes the deflection of the plate, the components of the *curvature* of the middle surface are defined by[3]

$$K_{\alpha\beta} = -w_{,\alpha\beta}. \tag{2. 3}$$

In the technical theory of plates, it is assumed that particles which, in the undeformed state, define a parallel to the x_3 axis, will after deformation define a normal to the bent middle surface. This assumption makes it possible to express the strain components in the plate in terms of $K_{\alpha\beta}$. Hooke's law then yields the stress components and the moments can be determined from (2. 1). Thus,[4]

$$M_{\alpha\beta} = D[(1-\nu)K_{\alpha\beta} + \nu K_{\gamma\gamma}\delta_{\alpha\beta}], \tag{2. 4}$$

where D is the flexural rigidity of the plate, ν is Poisson's ratio, and $\delta_{\alpha\beta}$ the Kronecker delta. Equation (2. 4) is easily solved for $K_{\alpha\beta}$. One finds

$$K_{\alpha\beta} = 1/D(1-\nu)[M_{\alpha\beta} - (\nu/1+\nu)M_{\gamma\gamma}\delta_{\alpha\beta}]. \tag{2. 5}$$

As is easily seen from (2. 3),

$$\epsilon_{\beta\gamma} K_{\alpha\beta,\gamma} = 0, \tag{2. 6}$$

where $\epsilon_{11} = \epsilon_{22} = 0$, $\epsilon_{12} = -\epsilon_{21} = 1$. Substitution of (2. 5) into (2. 6) yields

$$\epsilon_{\beta\gamma} M_{\alpha\beta,\gamma} - (\nu/1+\nu)\epsilon_{\alpha\beta} M_{\gamma\gamma,\beta} = 0. \tag{2. 7}$$

In the following, this equation will be called the *equation of compatibility*.

---

[2] In S. Timoshenko, *Theory of plates and shells* (1940), pp. 86-87 the moments are denoted by M_x, M_y, M_{xy} and M_{yx}. If the x and y axes coincide with the axes of x_1 and x_2, respectively, $M_x = M_{11}$, $M_y = M_{22}$, $M_{xy} = -M_{12}$ and $M_{yx} = M_{21}$. Thus, $M_{xy} = -M_{yx}$ but $M_{12} = M_{21}$.

[3] *Cf.* Timoshenko, *loc. cit.*, p. 36, where $r_x = 1/K_{11}$, $r_y = 1/K_{22}$, $r_{xy} = -1/K_{12}$.

[4] *Cf.* Timoshenko, *loc. cit.*, p. 88, Eqs. (99), (100).

3. Elastic plates and function space. Statement of the problem. In the following the term *state of stress* (or, more briefly, *state*) will be used to denote a set of continuous functions M_{11}, M_{22}, $M_{12} = M_{21}$. These functions are defined over the region R which is bounded by the contour of the plate. A state of stress may be conceived as representing a *point* in the function space defined by all states. Alternatively, a state may be considered as representing a *vector* $\boldsymbol{S}$ joining the origin to the point which is represented by this state.

Given a state $\boldsymbol{S}$, i. e. the three functions $M_{\alpha\beta}$, we may compute the corresponding components of curvature $K_{\alpha\beta}$ from (2. 5). The scalar product of any two states $\boldsymbol{S}$ and $\boldsymbol{S}'$ is defined by

$$\text{(3. 1)} \qquad \boldsymbol{S} \cdot \boldsymbol{S}' = \tfrac{1}{2} \int M_{\alpha\beta} K'_{\alpha\beta} dA,$$

where dA is the element of area of the middle surface of the plate and the integration is extended over the region R covered by the plate. On account of (2. 5),

$$\text{(3. 2)} \qquad \boldsymbol{S} \cdot \boldsymbol{S}' = 1/2D(1-\nu) \int [M_{\alpha\beta} M'_{\alpha\beta} - (\nu/1+\nu) M_{\alpha\alpha} M'_{\beta\beta}] dA.$$

The integrand on the right hand side is symmetric in $M_{\alpha\beta}$ and $M'_{\alpha\beta}$; accordingly

$$\text{(3. 3)} \qquad \boldsymbol{S} \cdot \boldsymbol{S}' = \boldsymbol{S}' \cdot \boldsymbol{S}.$$

The length $S = |\boldsymbol{S}|$ of a vector $\boldsymbol{S}$ is defined by

$$\text{(3. 4)} \qquad S^2 = \boldsymbol{S} \cdot \boldsymbol{S} = \tfrac{1}{2} \int M_{\alpha\beta} K_{\alpha\beta} dA$$
$$= 1/2D(1-\nu) \int [M_{\alpha\beta} M_{\alpha\beta} - (\nu/1+\nu) M_{\alpha\alpha} M_{\beta\beta}] dA.$$

The square of the length of $\boldsymbol{S}$ is thus seen to be the strain energy associated with the state $\boldsymbol{S}$. Since the integrand in (3. 4) is a positive definite form, $\boldsymbol{S} \cdot \boldsymbol{S}$ is never negative and vanishes if, and only if, all of the $M_{\alpha\beta}$ are zero.

The sum $\boldsymbol{S} + \boldsymbol{S}'$ of two states is defined as the state with the moments $M_{\alpha\beta} + M'_{\alpha\beta}$. The product of a state $\boldsymbol{S}$ and scalar quantity λ is defined as the state with the moments $\lambda M_{\alpha\beta}$. We note that

$$\boldsymbol{S} \cdot \boldsymbol{S}' = \boldsymbol{S}' \cdot \boldsymbol{S}$$
$$\boldsymbol{S} + \boldsymbol{S}' = \boldsymbol{S}' + \boldsymbol{S}$$
$$\boldsymbol{S} \cdot (\boldsymbol{S}'' + \boldsymbol{S}') = \boldsymbol{S} \cdot \boldsymbol{S}'' + \boldsymbol{S} \cdot \boldsymbol{S}'.$$

The *distance* between two states $\boldsymbol{S}$ and $\boldsymbol{S}'$ is defined by $|\boldsymbol{S} - \boldsymbol{S}'|$, where

$$\text{(3. 5)} \qquad |\boldsymbol{S} - \boldsymbol{S}'|^2 = \tfrac{1}{2} \int (M_{\alpha\beta} - M'_{\alpha\beta})(K_{\alpha\beta} - K'_{\alpha\beta}) dA.$$

Returning now to the problem of determining the bending and twisting moments in a plate which is subjected to the lateral load $p(x_1, x_2)$, we shall assume that the plate is clamped along the portion C_1 of its boundary and simply supported along the remainder C_2. The notations $M_{\alpha\beta}$ and $\boldsymbol{S}$ will be reserved for the so-called *natural* state which constitutes the solution of our problem. In addition to the boundary conditions on C_1 and C_2, the natural state satisfies the condition of equilibrium (2.2) and the condition of compatibility (2.7). We propose to approximate the natural state by *artificial* states for which some one or the other of these conditions are relaxed.

Let $\boldsymbol{S}^*$ be a state satisfying the equilibrium condition (2.2), and the boundary condition that the bending moment vanishes along C_2, i. e.,

$$M^*_{\alpha\beta} n_\alpha n_\beta = 0 \text{ along } C_2, \tag{3.6}$$

where n_α is the unit vector of the exterior normal. In accordance with the terminology introduced by Prager and Synge, such a state will be called a *completely associated state.* In addition to the state $\boldsymbol{S}^*$, we shall use a sequence of states $\boldsymbol{S}'_1, \boldsymbol{S}'_2, \cdots$ which will be called *homogeneous associated states*; the moments $M^{(p)\prime}{}_{\alpha\beta}$ of each state S'_p $(p = 1, 2, \cdots)$ satisfy the *homogeneous equilibrium condition*

$$M^{(p)\prime}{}_{\alpha\beta,\alpha\beta} = 0 \tag{3.7}$$

and the boundary condition

$$M^{(p)\prime}{}_{\alpha\beta} n_\alpha n_\beta = 0 \text{ along } C_2. \tag{3.8}$$

While equation (2.5) will yield the components of curvature corresponding to the states $\boldsymbol{S}^*$ and $\boldsymbol{S}'_p$, these need not satisfy (2.6). As a rule, it is therefore not possible to associate a deflection function $w(x_1, x_2)$ with these states.

In addition to the states discussed so far, we shall use a sequence of states $\boldsymbol{S}''_1, \boldsymbol{S}''_2, \cdots$ which satisfy the condition of compatibility and certain boundary conditions along C_1 and C_2. To construct these so-called *complementary states*, we start from deflection functions $w''_1, w''_2, \cdots$ which satisfy the boundary conditions

$$\left.\begin{array}{l} w''_q = 0 \text{ along } C_1 \text{ and } C_2 \\ \partial w''_q/\partial n = 0 \text{ along } C_1 \end{array}\right\} \quad (q = 1, 2, \cdots). \tag{3.9}$$

We compute the curvature components $K^{(q)\prime\prime}{}_{\alpha\beta}$ associated with these deflection functions from (2.3) and then the moments $M^{(q)\prime\prime}{}_{\alpha\beta}$ from (2.4). These

moments will satisfy the compatibility condition, but not, in general the equilibrium condition.

4. The hypercircle Γ. Let us consider the scalar product of the natural state $\boldsymbol{S}$ with itself. We have

$$(4.1)\qquad \boldsymbol{S}\cdot\boldsymbol{S}=\tfrac{1}{2}\int M_{\alpha\beta}K_{\alpha\beta}dA=-\tfrac{1}{2}\int M_{\alpha\beta}w_{,\alpha\beta}dA.$$

The last integral can be transformed as follows:

$$\begin{aligned}\int M_{\alpha\beta}w_{,\alpha\beta}dA &= \int M_{\alpha\beta}w_{,\alpha}n_{\beta}ds-\int M_{\alpha\beta,\beta}w_{,\alpha}dA\\ &=\int M_{\alpha\beta}w_{,\alpha}n_{\beta}ds-\int M_{\alpha\beta,\beta}wn_{\alpha}ds+\int M_{\alpha\beta,\alpha\beta}w\,dA.\end{aligned}$$

Here the second integral in the last line vanishes because $w=0$ along the entire boundary. Furthermore, the third integral can be written in the form $-\int p\,w\,dA$, on account of the equilibrium condition. The integrand of the first integral, finally, is a scalar which is most easily evaluated by choosing the coordinate axes parallel to the exterior normal and the tangent of the boundary. Thus, $n_1=1$, $n_2=0$, and

$$M_{\alpha\beta}w_{,\alpha}n_{\beta}=M_{11}w_{,1}+M_{21}w_{,2},$$

where $M_{11}=M_{\alpha\beta}n_{\alpha}n_{\beta}$ is the bending moment, $w_{,1}=\partial w/\partial n=w_{,\gamma}n_{\gamma}$ is the normal derivative and $w_{,2}=\partial w/\partial s$ is the tangential derivative of the deflection. As is easily seen by integrating by parts, the integral of the second term on the right-hand side along the closed contour vanishes. Accordingly,

$$\int M_{\alpha\beta}w_{,\alpha}n_{\beta}dA=\int(M_{\alpha\beta}n_{\alpha}n_{\beta})(w_{,\gamma}n_{\gamma})\,ds=0$$

since the first factor of the integrand vanishes along C_2 and the second along C_1. Thus,

$$(4.2)\qquad \boldsymbol{S}\cdot\boldsymbol{S}=\tfrac{1}{2}\int p\,w\,dA.$$

In a similar manner, it can be shown that

$$(4.3)\qquad \boldsymbol{S}^{*}\cdot\boldsymbol{S}=\tfrac{1}{2}\int p\,w\,dA,$$

$$(4.4)\qquad \boldsymbol{S}'_{p}\cdot\boldsymbol{S}=0,$$

$$(4.5)\qquad \boldsymbol{S}\cdot\boldsymbol{S}''_{q}=\tfrac{1}{2}\int p\,w''_{q}\,dA,$$

$$(4.6)\qquad \boldsymbol{S}^{*}\cdot\boldsymbol{S}''_{q}=\tfrac{1}{2}\int p\,w''_{q}\,dA,$$

$$(4.7)\qquad \boldsymbol{S}'_{p}\cdot\boldsymbol{S}''_{q}=0.$$

Combining equations (4.2) and (4.3), we see that

$$S \cdot (S - S^*) = 0 \tag{4.8}$$

or

$$(S - \tfrac{1}{2}S^*)^2 = \tfrac{1}{4}S^{*2}. \tag{4.9}$$

This shows that the endpoint of the vector S lies on the *hypersphere* with the center $\frac{1}{2}S^*$ and the radius $\frac{1}{2} \mid S^* \mid$.

Combining equations (4.5) and (4.6), we find, for $q = 1$,

$$S''_1 \cdot (S - S^*) = 0. \tag{4.10}$$

This shows that S and S^* have the same orthogonal projection on S''_1 and, hence, locates the endpoint of S on a certain hyperplane P_1 which is orthogonal to S''_1 and passes through the endpoint of S^*.

Since the endpoint of the vector S lies on the hypersphere (4.8) and on the hyperplane (4.10), it must lie on their intersection which will be called the hypercircle Γ_1. The center and radius of this hypercircle can be determined as follows. Let Σ_1 be the vector from the origin to the foot of the perpendicular dropped from the endpoint of S^* on the vector S''_1:

$$\Sigma_1 = S''_1(S^* \cdot S''_1/S''^2_1). \tag{4.11}$$

Obviously, Σ_1 is orthogonal to $\Sigma_1 - S^*$, i. e.,

$$\Sigma_1 \cdot (\Sigma_1 - S^*) = 0. \tag{4.12}$$

The endpoint of Σ_1 lies, therefore, on the hypersphere (4.8). It also lies on the hyperplane (4.10) and, hence, on the hypercircle Γ_1. Moreover, as is easily verified,

$$(S - \Sigma_1) \cdot (S - S^*) = 0. \tag{4.13}$$

Since the endpoints of S, Σ_1 and S^* all lie on the hypercircle Γ_1, equation (4.13) shows that the endpoints of Σ_1 and S^* are the endpoints of a diameter of Γ_1. The center of Γ_1 is therefore given by

$$C_1 = \tfrac{1}{2}(\Sigma_1 + S^*) = \tfrac{1}{2}[S^* + I''_1(S^* \cdot I''_1)], \tag{4.14}$$

where $I''_1 = S''_1/\mid S''_1 \mid$ is the unit vector which has the direction of S''_1. The radius R_1 of Γ_1 is given by

$$R_1^2 = (S^* - C)^2 = \tfrac{1}{4}[S^{*2} - (S^* \cdot I''_1)^2]. \tag{4.15}$$

5. Successive approximations to the moments $M_{\alpha\beta}$. In order to make R_1 vanish, Σ_1 and S^* must coincide. Now Σ_1, derived from S''_1, is derived from a deflection function w''_1, which satisfies the boundary conditions imposed on deflections; moreover, being derived from a deflection function,

S''_1 satisfies the condition of compatibility. The state S^*, on the other hand, satisfies the condition of equilibrium and the boundary conditions imposed on moments. When Σ_1 and S^* coincide, they therefore must both coincide with the natural state S which satisfies all conditions enumerated above. The statement $R_1 = 0$ thus is identical with the statement $C_1 = S$.

A good approximation C_1 to the natural state is therefore obtained when S^* and S''_1 are chosen so that R_1 becomes small. Now, the states S'_p and S''_q have been defined in such a manner that

$$S^* + \sum_{p=1}^{m} a_p S'_p$$

and

$$\sum_{q=1}^{n} b_q S''_q$$

may be used instead of S^* and S''_1 in defining a hypersphere and a hyperplane as loci for the endpoint of S. Given m states S'_p and n states S''_q, we can determine the coefficients a_p and b_q so that the radius of the corresponding hypercircle is minimized, and use the center of this hypercircle as the "best" approximation to the natural state.

In carrying out this program, we shall find it convenient to construct a sequence of *orthonormal* states I'_p from the given states S'_p and a sequence of orthonormal states I''_q from the given states S''_q the orthonormalization conditions being expressed by the equations

$$I'_r \cdot I'_s = \delta_{rs}, \tag{5.1}$$

$$I''_r \cdot I''_s = \delta_{rs}. \tag{5.2}$$

It is then found that a diameter of the "best" hypercircle $\Gamma_{m,n}$ is determined by the endpoints of the vectors

$$V^*_m = S^* - \sum_{p=1}^{m} I'_p (S^* \cdot I'_p) \tag{5.3}$$

$$V''_n = \sum_{q=1}^{n} I''_q (S^* \cdot I''_q). \tag{5.4}$$

The center $C_{m,n}$ of this best hypercircle is given by

$$\begin{aligned} C_{m,n} &= \tfrac{1}{2}(V^*_m + V''_n) \\ &= \tfrac{1}{2}[S^* - \sum_{p=1}^{m} I'_p (S^* \cdot I'_p) + \sum_{q=1}^{n} I''_q (S^* \cdot I''_q)] \end{aligned} \tag{5.5}$$

and its radius $R_{m,n}$ by

$$(5.6)\qquad R^2_{m,n} = (\boldsymbol{V}^*_m - \boldsymbol{C}_{m,n})^2 = \tfrac{1}{4}[S^{*2} - \sum_{p=1}^{m} (\boldsymbol{S}^* \cdot \boldsymbol{I}'_p)^2 - \sum_{q=1}^{n} (\boldsymbol{S}^* \cdot \boldsymbol{I}''_q)^2].$$

Equation (5.6) shows that $R_{m,n}$ can not increase as m or n is increased. By choosing the states $\boldsymbol{S}'_p$ and $\boldsymbol{S}''_q$ judiciously, we are able to arrive at rather small values of R for moderate values of m and n.

6. Application to the example of a clamped square plate. As a numerical example, we shall treat the problem of determining bending moments in a uniformly loaded square plate clamped along its edges. For convenience, we choose the intensity p of the uniformly distributed load to be unity, assume the elastic constants to have the values $\nu = 1/3$, $D = 1$, and let the vertices of the square be at the points $(1,1)$, $(1,-1)$, $(-1,-1)$, $(-1,1)$ in the x_1, x_2-plane.

For this problem the equation of equilibrium becomes

$$(6.1)\qquad M_{\alpha\beta,\alpha\beta} + 1 = 0,$$

while on the boundary of the square the conditions to be satisfied are

$$(6.2)\qquad w = 0, \quad \partial w/\partial n = 0.$$

Consequently the bending moments $M^*_{\alpha\beta}$ of the completely associated state $\boldsymbol{S}^*$ must satisfy (6.1), the bending moments $M^{(p)\prime}_{\alpha\beta}$ of the homogeneous associated states $\boldsymbol{S}'_p$ $(p = 1, 2, \cdots)$ must satisfy the homogeneous equation

$$(6.3)\qquad M^{(p)\prime}_{\alpha\beta,\alpha\beta} = 0,$$

and the complementary states $\boldsymbol{S}''_q$ $(q = 1, 2, \cdots)$ must be constructed from deflection functions w''_q satisfying the boundary conditions (6.2) along the edges of the plate.

The bending moments $M^{(p)\prime}_{\alpha\beta}$ for the states $\boldsymbol{S}'_p$ are derived from arbitrary stress functions $f_1^{(p)}$ and $f_2^{(p)}$ by means of the equations

$$(6.4)\qquad \begin{aligned} M^{(p)\prime}_{11} &= 2(\partial f^{(p)}_2/\partial x_2),\\ M^{(p)\prime}_{12} &= -(\partial f_1^{(p)}/\partial x_2 + \partial f_2^{(p)}/\partial x_1),\\ M^{(p)\prime}_{22} &= 2(\partial f^{(p)}_1/\partial x_1). \end{aligned}$$

When defined in the above manner, the bending moments $M^{(p)\prime}_{\alpha\beta}$ will satisfy (6.3) quite independently of how the functions $f^{(p)}_1$ and $f_2^{(p)}$ are chosen. Table 1 [5] lists the five homogeneous associated states $\boldsymbol{S}'_1, \cdots, \boldsymbol{S}'_5$ which we

[5] The bulk of the computations involved in the compilation of the tables presented in this section were carried through by Mr. C. C. Miesse.

use, giving in each case the functions $f^{(p)}{}_1$, $f^{(p)}{}_2$, the bending moments computed by means of (6.4) and the curvatures $K^{(p)\prime}{}_{\alpha\beta}$ computed using (2.5).

The bending moments $M^*{}_{\alpha\beta}$ for the state S^* are chosen to be those occurring in a circular clamped plate under a uniform load of unit intensity, since these moments are known and certainly satisfy (6.1). With proper choice of the radius of the circle, S^* would seem to be a reasonable first approximation to the solution S.[6] Incidentally, since these moments correspond to an actual deflection function, they satisfy the compatibility condition,

TABLE 1: The completely associated states S'_p.

$$M^{(p)\prime}{}_{11} = 2(\partial f^{(p)}{}_2/\partial x_2), \qquad K^{(p)\prime}{}_{11} = 3/8(3M^{(p)\prime}{}_{11} - M^{(p)\prime}{}_{22}),$$
$$M^{(p)\prime}{}_{12} = -(\partial f^{(p)}{}_1/\partial x_2 + \partial f^{(p)}{}_2/\partial x_1), \quad K^{(p)\prime}{}_{12} = 3/2M^{(p)\prime}{}_{12},$$
$$M^{(p)\prime}{}_{22} = 2(\partial f^{(p)}{}_1/\partial x_1), \qquad K^{(p)\prime}{}_{22} = 3/8(3M^{(p)\prime}{}_{22} - M^{(p)\prime}{}_{11})$$

S'_p	$f^{(p)}{}_1$	$f_2{}^{(p)}$	$M^{(p)\prime}{}_{11}$	$M^{(p)\prime}{}_{12}$	$M^{(p)\prime}{}_{22}$
S'_1	x_1	x_2	2	0	2
S'_2	$x_1x_2{}^2$	$x_2x_1{}^2$	$2x_1{}^2$	$-4x_1x_2$	$2x_2{}^2$
S'_3	$x_1x_2{}^4$	$x_2x_1{}^4$	$2x_1{}^4$	$-4(x_1x_2{}^3+x_1{}^3x_2)$	$2x_2{}^4$
S'_4	$x_1{}^3$	$x_2{}^3$	$6x_2{}^2$	0	$6x_1{}^2$
S'_5	$x_1{}^3x_2{}^2$	$x_2{}^3x_1{}^2$	$6x_1{}^2x_2{}^2$	$-2(x_1{}^3x_2+x_2{}^3x_1)$	$6x_1{}^2x_2{}^2$

S'_p	$K^{(p)\prime}{}_{11}$	$K^{(p)\prime}{}_{12}$	$K^{(p)\prime}{}_{22}$
S'_1	$+3/2$	0	$+3/2$
S'_2	$+3/4(3x_1{}^2-x_2{}^2)$	$-6x_1x_2$	$+3/4(3x_2{}^2-x_1{}^2)$
S'_3	$+3/4(3x_1{}^4-x_2{}^4)$	$-6(x_1x_2{}^3+x_2x_1{}^3)$	$+3/4(3x_2{}^4-x_1{}^4)$
S'_4	$+9/4(3x_2{}^2-x_1{}^2)$	0	$+9/4(3x_1{}^2-x_2{}^2)$
S'_5	$+9/2x_1{}^2x_2{}^2$	$-3(x_1{}^3x_2+x_2{}^3x_1)$	$+9/2x_1{}^2x_2{}^2$

which is more than is actually required by the theory. The complementary states S''_1, S''_2, S''_3 which we use are listed in Table 2 where the deflection function, curvatures from (2.3) and bending moments from (2.4) are

[6] That value of a is chosen for which $S^* \cdot S^*$ is made a minimum, it being noted that $S^* \cdot S^*$ is the first term appearing in the error formula (5.6). The value obtained is $a^2 = 1.565217$.

tabulated for each. For convenience, the completely associated state S^* has been included in Table 2 with the complementary states [note that $a^2 = 1.565217$, (2. 3) yields the $K^*_{\alpha\beta}$ from w^*, (2. 4) yields the $M^*_{\alpha\beta}$ from the $K^*_{\alpha\beta}$].

Orthonormalizing the states $S'_1, \cdots, S'_5$ we obtain the states

$$I'_p = \sum_{i=1}^{p} a_{ip} S'_i, \qquad (p = 1, \cdots, 5), \tag{6. 5}$$

where the coefficients a_{ip} are determined below. In addition, the states S''_1, S''_2, S''_3 are orthonormalized to give the states

$$I''_q = \sum_{i=1}^{q} b_{iq} S''_i, \qquad (q = 1, 2, 3). \tag{6. 6}$$

With these notations, (5. 5) becomes

TABLE 2: The complementary states S''_q and the completely associated state S^*
(Note: $a^2 = 1.565217$)

S''_q	w''_q	$K^{(q)''}_{11} = -\partial^2 w''_q/\partial x_1^2$
S''_1	$(1-x_1^2)^2(1-x_2^2)^2$	$-4(1-x_2^2)^2(3x_1^2-1)$
S''_2	$(1-x_1^4)^2(1-x_2^4)^2$	$-8(1-x_2^4)^2(7x_1^6-3x_1^2)$
S''_3	$(1-x_1^6)^2(1-x_2^6)^2$	$-12(1-x_2^6)^2(11x_1^{10}-5x_1^4)$
S^*	$w^* = 1/64(a^2-x_1^2-x_2^2)^2$	$K^*_{11} = 1/16(a^2-3x_1^2-x_2^2)$

S''_q	$K^{(q)''}_{12} = -\partial^2 w''_q/\partial x_1 \partial x_2$	$K^{(q)''}_{22} = -\partial^2 w''_q/\partial x_2^2$
S''_1	$-16x_1x_2(1-x_1^2)(1-x_2^2)$	$-4(1-x_1^2)^2(3x_2^2-1)$
S''_2	$-64x_1^3x_2^3(1-x_1^4)(1-x_2^4)$	$-8(1-x_1^4)^2(7x_2^6-3x_2^2)$
S''_3	$-144(x_1^{11}-x_1^5)(x_2^{11}-x_2^5)$	$-12(1-x_1^6)^2(11x_2^{10}-5x_2^4)$
S^*	$K^*_{12} = -1/8x_1x_2$	$K^*_{22} = 1/16(a^2-3x_2^2-x_1^2)$

S''_q	$M^{(q)''}_{11} = 1/3(3K^{(q)''}_{11} + K^{(q)''}_{22})$	$M^{(q)''}_{12} = 2/3K^{(q)''}_{12}$
S''_1	$-4/3[3(1-x_2^2)^2(3x_1^2-1)$ $+ (1-x_1^2)^2(3x_2^2-1)]$	$-32/3x_1x_2(1-x_1^2)(1-x_2^2)$

S''_2	$-8/3[3(1-x_2^4)^2(7x_1^6-3x_1^2) + (1-x_1^4)^2(7x_2^6-3x_2^2)]$	$-128/3x_1^3x_2^3(1-x_1^4)(1-x_2^4)$
S''_3	$-4[3(1-x_2^6)^2(11x_1^{10}-5x_1^4) + (1-x_1^6)^2(11x_2^{10}-5x_2^4)]$	$-96(x_1^{11}-x_1^5)(x_2^{11}-x_2^5)$
S^*	$M^*_{11}=1/24(2a^2-5x_1^2-3x_2^2)$	$M^*_{12}=-1/12x_1x_2$

S''_q	$M^{(q)''}_{22}=1/3(3K^{(q)''}_{22}+K^{(q)''}_{11})$
S''_1	$-4/3[3(1-x_1^2)^2(3x_2^2-1)+(1-x_2^2)^2(3x_1^2-1)]$
S''_2	$-8/3[3(1-x_1^4)^2(7x_2^6-3x_2^2)+(1-x_2^4)^2(7x_1^6-3x_1^2)]$
S''_3	$-4[3(1-x_1^6)^2(11x_2^{10}-5x_2^4)+(1-x_2^6)^2(11x_1^{10}-5x_1^4)]$
S^*	$M^*_{22}=1/24(2a^2-5x_2^2-3x_1^2)$

$$C_{m,n}=\tfrac{1}{2}S^*-\tfrac{1}{2}\sum_{p-1}^{m}[\sum_{i=1}^{p}\sum_{j=1}^{p}a_{ip}a_{jp}(S^*\cdot S'_j)S'_i] + \tfrac{1}{2}\sum_{q-1}^{n}[\sum_{i=1}^{q}\sum_{j=1}^{q}b_{iq}b_{jq}(S^*-S''_j)S''_i], \tag{6.7}$$

where $C_{m,n}$ is the state which we choose as our best approximation to the solution S of the problem after performing m iterations on the associated states S'_p and n iterations on the complementary states S''_q. The error made with this approximation, given by the quantity $R_{m,n}$ of (5.6), takes the form

$$R^2_{m,n}=\tfrac{1}{4}S^*\cdot S^*-\tfrac{1}{4}\sum_{p=1}^{m}[\sum_{i=1}^{p}\sum_{j=1}^{p}a_{ip}a_{jp}(S^*\cdot S'_i)(S^*\cdot S'_j)] -\tfrac{1}{4}\sum_{q=1}^{n}[\sum_{i=1}^{q}\sum_{j=1}^{q}b_{iq}b_{jq}(S^*\cdot S''_i)(S^*\cdot S''_j)]. \tag{6.8}$$

We introduce the notations $r'_{ip}=a_{ip}/a_{pp}$ and $r''_{iq}=b_{iq}/b_{qq}$ [7] whence

$$\begin{aligned} a_{ip}a_{jp}&=r'_{ip}r'_{jp}a_{pp}^2,\\ b_{iq}b_{jq}&=r''_{iq}r''_{jq}b_{qq}^2. \end{aligned} \tag{6.9}$$

Substituting these relations in (6.7) and (6.8) we get

$$C_{m,n}=\tfrac{1}{2}S^*-\tfrac{1}{2}\sum_{p=1}^{m}a_{pp}^2[\sum_{i=1}^{p}\sum_{j=1}^{p}r'_{ip}r'_{jp}(S^*\cdot S'_j)S'_i] +\tfrac{1}{2}\sum_{q=1}^{n}b_{qq}^2[\sum_{i=1}^{q}\sum_{j=1}^{q}r''_{iq}r''_{jq}(S^*\cdot S''_j)S''_i] \tag{6.10}$$

[7] In this section summation signs are written explicitly; repeated indices do not of themselves imply summation.

$$(6.11)\quad R^2_{m,n}=\tfrac{1}{4}S^*\cdot S^*-\tfrac{1}{4}\sum_{p=1}^{m}a^2_{pp}\Big[\sum_{i=1}^{p}\sum_{j=1}^{p}r'_{ip}r'_{jp}(S^*\cdot S'_i)(S^*\cdot S'_j)\Big]$$
$$-\tfrac{1}{4}\sum_{q=1}^{n}b^2_{qq}\Big[\sum_{i=1}^{q}\sum_{j=1}^{q}r''_{iq}r''_{jq}(S^*\cdot S''_i)(S^*\cdot S''_j)\Big].$$

Following the well-known procedure, the constants r'_{ip} and r'_{iq} are found by solving simultaneous linear equations. Thus, for a fixed value of p, r'_{ip}, $(i=1,\cdots,p-1$; note $r'_{pp}=1)$ are the solutions of the $(p-1)$ equations

$$(6.12)\qquad \sum_{i=1}^{p-1}(S'_j\cdot S'_i)r'_{ip}=-S'_j\cdot S'_p,\qquad j=1,\cdots,p-1.$$

When the r'_{ip} have been found, the constant a^2_{pp} is given by

$$(6.13)\qquad a^2_{pp}=\Big[\sum_{i=1}^{p}r'_{ip}(S'_p\cdot S'_i)\Big]^{-1}.$$

The values of the scalar products $S'_i\cdot S'_j$ which appear as coefficients in (6.12) are tabulated in Table 3. The values obtained for the r'_{ip} as well as the values of the constants $a^2_{11},\cdots,a^2_{55}$ are to be found in Table 4. Similarly, the quantities r''_{iq} $(i=1,\cdots,q-1$; note $r''_{qq}=1)$ are found by solving the equations

$$(6.14)\qquad \sum_{i=1}^{q-1}(S''_j\cdot S''_i)r''_{iq}=-S''_j\cdot S''_q,\qquad j=1,\cdots q-1$$

and b^2_{qq} is then determined from the formula

$$(6.15)\qquad b^2_{qq}=\Big[\sum_{i=1}^{q}r''_{iq}(S''_q\cdot S''_i)\Big]^{-1}.$$

The values $S''_i\cdot S''_j$ are given in Table 5 and constants r''_{iq} and b^2_{11}, b^2_{22}, b^2_{33} are given in Table 6.

TABLE 3: $S'_i\cdot S'_j=\tfrac{1}{2}\int_{-1}^{1}\int_{-1}^{1}(M^{(i)\prime}_{11}K^{(j)\prime}_{11}+2M^{(i)\prime}_{12}K^{(j)\prime}_{12}+M^{(i)\prime}_{22}K^{(j)\prime}_{22})\,dx_1dx_2.$

	S'_1	S'_2	S'_3	S'_4	S'_5	S^*
S'_1	12.0	4.0	2.40	12.0	4.0	0.115942
S'_2	4.0	13.60	14.971429	2.40	8.80	0.127536
S'_3	2.40	14.971429	18.582857	1.028571	10.125714	0.175569
S'_4	12.0	2.40	1.028571	26.40	7.20	—0.17392
S'_5	4.0	8.80	10.125714	7.20	8.525714	—0.005797
S^*	0.115942	0.127536	0.175569	—0.173920	—0.005797	0.022417

TABLE 4: Values of r'_{ip} and a_{ii}^2.

	$p=1$	2	3	4	5	a_{ii}^2
$i=1$	1.000000	-0.333333	0.185093	-1.094489	0.104660	0.083333
2	....	1.000000	-1.155280	0.448824	-0.268820	0.081522
3	....	....	1.000000	$-.275592$	0.326164	0.577738
4	....	....	....	1.000000	-0.283154	0.071125
5	....	....	....	....	1.000000	0.808149

TABLE 5: $S''_i \cdot S''_j = \frac{1}{2} \int_{-1}^{1} \int_{-1}^{1} (M^{(i)''}_{11} K^{(j)''}_{11} + 2M^{(i)''}_{12} K^{(j)''}_{12} + M^{(i)''}_{22} K^{(j)''}_{22})\, dx_1 dx_2.$

	S''_1	S''_2	S''_3	S^*
S''_1	26.749384	38.955228	43.195622	0.568889
S''_2	38.955228	155.323820	257.190746	1.011358
S''_3	43.195622	257.190746	523.247578	1.252023
S^*	0.568889	1.011358	1.252023	0.022417

TABLE 6: Values of r''_{iq} and b_{ii}^2.

	$q=1$	2	3	b_{ii}^2
$i=1$	1.000000	-1.456304	1.254957	0.037384
2	....	1.000000	-1.970592	0.010143
3	....	....	1.000000	0.014157

All these constants being determined, we can use (6. 10) and (6. 11) to determine $C_{m,n}$ and $E^2_{m,n}$. We note that the values of the scalar products $S^* \cdot S'_i$ and $S^* \cdot S''_i$ are also required in order to do this; these values are listed in Table 3 and Table 5 respectively. If we so desire, we can now write $C_{m,n}$ in the form

$$C_{m,n} = \tfrac{1}{2} S^* + \alpha_1 S'_1 + \cdots + \alpha_m S'_m + \beta_1 S''_1 + \cdots + \beta_n S_n,$$

where the $\alpha_1, \cdots, \alpha_m, \beta_1, \cdots, \beta_n$ are known constants. If, then, we designate by $M_{\alpha\beta}{}^{(m,n)}$ the bending moments corresponding to the state $\boldsymbol{C}_{m,n}$ we have

$$M_{\alpha\beta}{}^{(m,n)} = \tfrac{1}{2}M^*{}_{\alpha\beta} + \alpha_1 M_{\alpha\beta}{}^{(1)\prime} + \cdots + \alpha_m M_{\alpha\beta}{}^{(m)\prime} + \beta_1 M_{\alpha\beta}{}^{(1)\prime\prime} + \cdots + \beta_n M_{\alpha\beta}{}^{(n)\prime\prime}.$$

The moments $M^*{}_{\alpha\beta}$, $M_{\alpha\beta}{}^{(1)\prime}, \cdots, M_{\alpha\beta}{}^{(5)\prime}$, $M_{\alpha\beta}{}^{(1)\prime\prime}$, $M_{\alpha\beta}{}^{(2)\prime\prime}$, $M_{\alpha\beta}{}^{(3)\prime\prime}$ are given in our tables and so we can explicitly write out the bending moment functions $M_{\alpha\beta}{}^{(m,n)}$ which constitute our approximations to the true bending moments $M_{\alpha\beta}$.

Actually it is awkward to write these formulas explicitly and so we have chosen simply to evaluate $M_{11}{}^{(m,n)}$ numerically at two convenient points in the plate, the center and midpoint of a side. This involves replacing $\boldsymbol{S}'_i$ in (6.10) by $M_{11}{}^{(i)\prime}(x_1, x_2)$ and $\boldsymbol{S}''_i$ by $M_{11}{}^{(i)\prime\prime}(x_1, x_2)$ where these functions, found in Tables 1 and 2 are to be evaluated at the point under consideration. Thus, we find approximate values for the bending moment per unit length in the x_2 direction at the points $(0, 0)$ and $(1, 0)$. The values obtained at the successive iterations are given in Table 7 below. Also given is the measure of error $R^2{}_{m,n}$, involved in taking $\boldsymbol{C}_{m,n}$ as an approximation to the solution $\boldsymbol{S}$. This error, when small indicates, as previously pointed out, that

TABLE 7: The bending moments $M_{11}(0, 0)$ and $M_{11}(1, 0)$ and error $R^2{}_{m,n}$ are determined at successive iterations.

$\boldsymbol{C}_{m,n}$	$M_{11}{}^{(m,n)}(0,0)$	$M_{11}{}^{(m,n)}(1,0)$	$R^2{}_{m,n}$
$\boldsymbol{C}_{1,1}$	0.112267	—0.133680	0.002300
$\boldsymbol{C}_{2,2}$	0.107480	—0.157387	0.002113
$\boldsymbol{C}_{3,3}$	0.100886	—0.154239	0.001533
$\boldsymbol{C}_{4,3}$	0.090343	—0.166510	0.001207
$\boldsymbol{C}_{5,3}$	0,083550	—0.196222	—0.000095*

the *overall* approximation of $\boldsymbol{C}_{m,n}$ to $\boldsymbol{S}$ is good. However, at any particular point, $R_{m,n}$ may not be taken as the error in the bending moment. It is only true that as $R_{m,n}$ decreases to zero, the approximation $\boldsymbol{C}_{m,n}$ will converge to $\boldsymbol{S}$ and the error at any point will approach zero. We note in the above table the decrease in the value of $R_{m,n}$ as successive iterations are made. The last entry, which is starred, came out negative in apparent contradiction to the positive character of $R^2{}_{m,n}$. However, by the fifth iteration, accumulated

errors are sufficiently great that an error of 1 in the fourth decimal place is not unlikely. Consequently, we may only say that to at least three decimal places $R^2{}_{5,5}$ vanishes.

The last iteration gives us the approximate values $M_{11}(0,0) = .084$, $M_{11}(1,0) = -.196$. These agree to within 5% with the values $M_{11}(0,0) = .0888$, $M_{11}(1,0) = -.2060$ found by Nádai.[8]

Conclusion. To obtain the accuracy desired for most engineering purposes a slightly larger number of steps should be carried out. The numerical processes encountered in the application of the present method are particularly well suited to modern automatic computing equipment.

BROWN UNIVERSITY.

[8] A. Nádai, *Die elastischen Platten*, Berlin, 1925, p. 184.

ON THE LINEAR LOGARITHMICO-EXPONENTIAL DIFFERENTIAL EQUATION OF THE SECOND ORDER.*

By PHILIP HARTMAN

1. This paper is concerned with the asymptotic behavior, for large t-values, of the solutions $x = x(t)$ of the differential equation

$$(1) \qquad d^2x/dt^2 + f(t)x = 0,$$

where $f = f(t)$ is a logarithmico-exponential (L-function) in the sense of Hardy [3], p. 17. Since $f(t)$ is an L-function,

$$(2) \qquad f(t) \text{ is monotone for large } t\text{-values}$$

and

$$(3) \qquad \lim_{t\to\infty} f(t) = \lambda \text{ exists,}$$

if $\lambda = \pm\infty$ is permitted in (3). The statements to be proved depend on the following known results:

(i) if $\infty > \lambda > 0$, then (2) implies that (1) possesses a pair of (linearly independent) solutions of the respective asymptotic forms

$$(4_i) \qquad \cos\int^t |f(s)|^{\frac{1}{2}}\,ds + o(1), \qquad \sin\int^t |f(s)|^{\frac{1}{2}}\,ds + o(1)$$

as $t\to\infty$; [8], p. 270;

(ii) if $0 > \lambda > -\infty$, then (i) remains true if (4_i) is replaced by

$$(4_{ii}) \quad (1+o(1))\exp\int^t |f(s)|^{\frac{1}{2}}\,ds, \qquad (1+o(1))\exp\int^t -|f(s)|^{\frac{1}{2}}\,ds,$$

[10], p. 60: finally,

(iii) if

$$(5) \qquad \int^\infty t\,|f(t)|\,dt < \infty,$$

then (1) possesses a pair of (linearly independent) solutions of the respective asymptotic forms

$$(4_{iii}) \qquad 1 + o(1), \qquad (1+o(1))t;$$

* Received March 13, 1948

as pointed out in [10], p. 66, this result can be obtained from [1], § 4, by a suitable change of the independent variable; for a simple proof avoiding successive approximations, cf. [10].

The theorems (i)-(iii) fail to furnish asymptotic formulae for the solutions of (1) when the limit (3) satisfies $|\lambda| = \infty$ or when $\lambda = 0$ but (5) does not hold. It turns out however that even in these exceptional cases, the variables can be changed in such a manner that (1) is transformed into an equation to which one of the above theorems is applicable. The required change of variables consists of a finite number of repeated applications of a standard substitution (Liouville, Riemann). The applicability of this substitution in the study of the asymptotic behavior of solutions of (1) has been pointed out in [9].

2. The resulting asymptotic formulae for the solutions of (1) can be stated directly in terms of $f(t)$ as follows:

Let $l_k(t)$ denote the k-iterated logarithm of t, that is, $l_1(t) = \log t$, $l_2(t) = \log\log t, \cdots$; and for $n = 0, \cdots$, let

$$L_n(t) = \prod_{k=1}^{n} l_k^{-2}(t), \tag{6}$$

where, as usual, the empty product is

$$L_0(t) \equiv 1. \tag{7}$$

Hence,

$$L_{n+1}(t) = L_n(t)/l_{n+1}^2(t), \qquad (n = 0, 1, \cdots). \tag{8}$$

Thus, for any L-function $f(t)$ and any non-negative integer n,

$$\lim_{t\to\infty} \left(4t^2 f(t) - \sum_{j=0}^{n} L_j(t)\right)/L_n(t) = \lambda_n \tag{9_n}$$

exists, if $\lambda_n = \pm\infty$ is admitted. It will be verified below (Lemma 1) that $\lambda_n = 0$ cannot hold for $n = 0, 1, \cdots$, so that there exists a least non-negative integer $n = M = M(f)$ for which $\lambda_n \neq 0$,

$$\lambda_k = 0 \text{ for } k = 0, 1, \cdots, M-1 \text{ and } \lambda_M \neq 0, \qquad (M \geqq 0). \tag{10}$$

Define a finite or an infinite sequence of L-functions $F_n(t)$ for large t-values by induction as follows: Let

$$F_0(t) = f(t) \tag{11}$$

and suppose $F_0, \cdots, F_n$ have been defined for large t-values and are L-functions. Let

$$(12_n) \qquad P_n = P_n(t) = \prod_{k=0}^{n} | F_k(t) |^{\frac{1}{4}},$$

where, as before, the empty product is

$$P_{-1}(t) \equiv 1.$$

If

$$(13_n) \qquad \int^{\infty} P_n(t)\, dt < \infty,$$

then $F_{n+1}(t)$ remains undefined. On the other hand, if

$$(14_n) \qquad \int^{\infty} P_n(t)\, dt = \infty,$$

then the L-function $P_n(t)$ is not identically zero and, hence, does not vanish for the large t-values. In view of (12_n), the same is true of the L-function $F_n(t)$. In this case, $F_{n+1}(t)$ is defined by

$$(15_n) \quad F_{n+1}(t) = \pm 1 + | F_n |^{-3/4} P_{n-1}^{-1} \frac{d}{dt} (P_{n-1}^{-1} \frac{d}{dt} | F_n |^{-1/4}),$$

where $\pm$ is the (constant) algebraic sign of $F_n(t)$ for large t-values.

Finally, let $N = N(f)$ be the non-negative integer which is the minimum of $M(f)$ and the least non-negative integer n, if any, for which (13_n) holds. Thus, $0 \leqq N \leqq M$; also,

$$(16) \qquad \int^{\infty} P_N(t)\, dt < \infty$$

holds if $N < M$, but may or may not hold if $N = M$; finally, if $N > 0$, then (14_n) holds for $n = 0, \cdots, N-1$.

It will be proved that if (16) holds, then (1) possesses a pair of (linearly independent) solutions of the respective asymptotic forms (4_{iii}) or of the respective asymptotic forms

$$(17) \quad (1 + o(1))(tl_1(t) \cdots l_{N-1}(t))^{\frac{1}{2}}, \quad (1 + o(1))(tl_1(t) \cdots l_{N-1}(t))^{\frac{1}{2}} l_N(t)$$

according as $N = 0$ or $N > 0$.

If (16) does not hold, so that $M = N$, the situation is as follows: If $\infty \geqq \lambda_N > 0$, then (1) possesses a pair of (linearly independent) solutions of the respective asymptotic forms

$$(18) \quad P_{N-1}^{-\frac{1}{2}}(t)\{\cos \int^t P_N(s)\, ds + o(1)\},$$

$$P_{N-1}^{-\frac{1}{2}}(t)\{\sin \int^t P_N(s)\, ds + o(1)\},$$

where the factor $P_{N-1}^{-\frac{1}{2}}(t)$ is asymptotically proportional to

(19) $$f^{-1/4}(t) \text{ or } (tl_1(t)\cdots l_{N-1}(t))^{\frac{1}{2}}$$

according as $N \leqq 1$ or $N > 1$. If $0 > \lambda_N \geqq -\infty$, then (1) possesses a pair of (linearly independent) solutions of the respective asymptotic forms

(20) $$(1+o(1))P_N^{-\frac{1}{2}}(t)\exp\int^t P_N(s)ds,$$
$$(1+o(1))P_N^{-\frac{1}{2}}(t)\exp\int^t -P_N(s)ds.$$

3. The proof of these assertions will depend on a number of lemmas which do not involve differential equations but deal only with the properties of L-functions. The lemmas will be stated now, but the proofs will be postponed until **5-9**. It is supposed throughout that $f(t)$ is an L-function.

LEMMA 1. *There exists a least non-negative integer $n = M = M(f)$ for which the limit (9_n) is not zero.*

LEMMA 2. *If the limit (3) satisfies $0 < \lambda \mid \leqq \infty$, then $F_1(t)$, defined by (15_0) for large t-values, satisfies*

(21) $$\lim_{t\to\infty} F_1(t) = \pm 1,$$

where $\pm 1 = \operatorname{sgn} f(t)$ for large t; in fact,

(22) $$\int^\infty |\pm 1 - F_1(t)|\,|f(t)|^{\frac{1}{2}}dt < \infty.$$

LEMMA 3. *If $F_0, F_1, \cdots, F_n$ are defined by (11) and (15_n) for large t-values (so that (14_k) holds for $k = 0, 1, \cdots, n-1$), then*

(23_n) $$\lim_{t\to\infty} F_k(t) = 0 \text{ for } k = 1, \cdots, n$$

holds if and only if $M(f) \geqq n$, that is, the limits (9_k) satisfy

(24_n) $$\lambda_k = 0 \text{ for } k = 1, \cdots, n-1.$$

LEMMA 4. *If $N(f) > 1$, the functions $F_0, F_1, \cdots, F_{N-1}$ satisfy*

(25_0) $$F_0(t) \sim 1/4t^2$$

and

(25_n) $$F_n(t) \sim l_n^{-2}(t), \qquad (0 < n < N).$$

If $N(f) > 0$, then

(26) $$\int^t P_{N-1}(t)\,dt \sim \tfrac{1}{2}l_N(t), \qquad (0 < N).$$

LEMMA 5. *If* (16) *does not hold* (*so that* $M(f) = N(f)$), *then*

$$\lim_{t\to\infty} F_N(t) = \lambda_N/|1+\lambda_N| \neq 0 \tag{27}$$

(*where it is understood that*

$c/|1+c|$ denotes -1, $-\infty$ or 1 if $c = -\infty$, -1 or $+\infty$,

respectively).

4. These lemmas will be assumed for the moment and will be used to prove the assertions made at the end of **2** concerning the asymptotic behavior of solutions of (1). In order to systematize the notations, let

$$x_0 = x \text{ and } t_0 = t, \tag{28}$$

so that, by (11), equation (1) becomes

$$d^2x_0/dt_0^2 + F_0(t_0)x_0 = 0. \tag{29}$$

If $F_0 \not\equiv 0$ and if the new variables (Liouville, Riemann)

$$t_1 = t_1(t_0) = \int^{t_0} |F_0(t_0)|^{\frac{1}{2}}\, dt_0 \tag{30}$$

and

$$x_1 = x_1(t_1) = x_0(t_0(t_1))\,|F_0(t_0(t_1))|^{1/4} \tag{31}$$

are introduced, (29) is transformed into

$$d^2x_1/dt_1^2 + F_1(t_0(t_1))x_1 = 0, \tag{32}$$

where $t_0 = t_0(t_1)$ is the inverse of (30). It is, of course, understood that the lower limit of integration in (30) is fixed so large that $F_0(t_0)$ does not vanish for any greater value of t_0.

If the pairs of variables $x_0, t_0; x_1, t_1; \cdots; x_{n-1}, t_{n-1}$ have been defined and $n \leq N$, let

$$t_n = t_n(t_{n-1}) = \int^{t_{n-1}} |F_{n-1}(t_0(t_{n-1}))|^{\frac{1}{2}}\, dt_{n-1}, \tag{33}$$

that is,

$$t_n = t_n(t_0) = \int^{t_0} P_{n-1}(t_0)\, dt_0, \tag{34_n}$$

and

$$x_n = x_n(t_n) = x_{n-1}(t_{n-1}(t_n))\,|F_{n-1}(t_0(t_{n-1}))|^{1/4}. \tag{35_n}$$

The equation (1) is transformed into

$$d^2x_n/dt_n^2 + F_n(t_0(t_n))x_n = 0. \tag{36_n}$$

In virtue of the definition of $N(f)$ and (34_n),

(37) $$t_n(t_0) \to \infty \text{ as } t_0 \to \infty \qquad (n \leqq N).$$

Suppose that (16) holds. In view of (33) and (34_n), the inequality (16) implies

(38) $$\int^{\infty} | F_n(t_0(t_N)) |^{\frac{1}{2}} dt_N < \infty.$$

From the monotony of F_N, it follows that

$$t_N | F_N(t_0(t_N)) |^{\frac{1}{2}} \to 0 \text{ as } t_N \to \infty.$$

Hence, as $t_N \to \infty$,

$$t_N | F_N(t_0(t_N)) | = o(1) | F_N(t_0(t_N)) |^{\frac{1}{2}},$$

so that (38) implies

(39) $$\int^{\infty} t_N | F_N(t_0(t_N)) | dt_N < \infty.$$

Consequently, the theorem (iii) quoted in **1** shows that (36_N) possesses a pair of (linearly independent) solutions $x_N = x_N(t_N)$ of the respective asymptotic forms

$$1 + o(1), \qquad (1 + o(1))t_N,$$

as $t_N \to \infty$. In virtue of (34_N), (25_N) and (37), it follows that (1) possesses a pair of linearly independent solutions of the respective asymptotic forms

$$(1 + o(1))P_{N-1}^{-\frac{1}{2}}(t), \qquad (1 + o(1))P_{N-1}^{-\frac{1}{2}}(t) \int^{t} P_{N-1}(t)dt.$$

Hence, (12_n) and Lemma 4 show that (1) possesses a pair of linearly independent solutions of the respective forms (4_{iii}) or (17) according as $N = 0$ or $N > 0$.

Suppose now that (16) does not hold, so that $M = N$. Then by Lemma 5 and (37)

(40) $$\lim_{t \to \infty} F_N(t_0(t_N)) = \lambda_N / | 1 + \lambda_N |.$$

Suppose that

(41) $$0 < | \lambda_N | / | 1 + \lambda_N | < \infty.$$

Then it follows from the theorems (i) and (ii) quoted in **1**, that (36_N) has a pair of (linearly independent) solutions $x_N = x_N(t_N)$ of either the respective asymptotic forms

$$\cos \int^{t_N} | F_N(t_0(t_N)) |^{\frac{1}{2}} dt_N + o(1), \qquad \sin \int^{t_N} | F_N(t_0(t_N)) |^{\frac{1}{2}} dt_N + o(1)$$

or of the respective asymptotic forms

$$(1+o(1))\exp\int^{t_N}|F_N(t_0(t_N))|^{\frac{1}{2}}\,dt_N, \qquad (1+o(1))\exp\int^{t_N}-|F_N(t_0(t_N))|^{\frac{1}{2}}$$

according as $\lambda_N>0$ or $\lambda_N<0$. In virtue of (34_N), (35_N) and (37), it follows that (1) possesses a pair of (linearly independent) solutions of either the respective asymptotic forms (18) or of the forms (20) according as $\lambda_N>0$ or $\lambda_N<0$, when (41) holds. (It would seem that the factor in (20) comes out to be $P_{N-1}^{-\frac{1}{2}}$ instead of $P_N^{-\frac{1}{2}}$, but since the limit (40) is finite and non-zero and since a constant multiple of a solution of (1) is again a solution, (20) is valid.) The statement concerning $P_{N-1}^{-\frac{1}{2}}(t)$ following (18) is a consequence of Lemma 4.

It remains to remove the hypothesis (41). Actually, the first inequality in (41) must hold, so that only the second inequality is an assumption. The formula line following (27) shows that of the two possibilities, $\pm\infty$, the only one that can occur is that (40) is $-\infty$ (corresponding to $\lambda_N=-1$). If the limit (40) is $-\infty$, it follows from (33) that

$$t_{N+1}(t_N)\to\infty \text{ as } t_N\to\infty$$

and from Lemma 2 that

$$\lim_{t_{N+1}\to\infty}F_{N+1}(t_0(t_{N+1}))=-1,$$

where $-1=\operatorname{sgn}\lambda_N$; and, in fact, from (22),

$$\int^{\infty}|-1-F_{N+1}(t_0(t_{N+1}))|\,dt_{N+1}<\infty. \tag{42}$$

It is known that (42) implies that (36_{N+1}) possesses a pair of (linearly independent) solutions $x_{N+1}=x_{N+1}(t_{N+1})$ of the respective asymptotic forms

$$(1+o(1))\exp t_{N+1}, \qquad (1+o(1))\exp(-t_{N+1}).$$

This result can be deduced from [1], § 2-§ 3, by a suitable change of variables; cf. also [10], p. 66 and [4], § 18. Again, it follows that (1) possesses a pair of (linearly independent) solutions of the respective asymptotic forms (20).

This completes the proof of the assertions at the end of **2**. It remains to prove the lemmas of **3** on which the proof is based.

5. Proof of Lemma 1. The definition (9_n) of λ_n shows that $\lambda_{n+1}=0$ means that

$$4t^2f(t)-\sum_{j=0}^{n}L_j(t)\sim L_{n+1}(t)$$

or, in view of (8),

$$(4t^2f(t)-\sum_{j=0}^{n}L_j(t))/L_n(t)\sim 1/l_{n+1}{}^2(t). \tag{43}$$

In order to prove the Lemma 1, it is sufficient to show that this relation cannot hold for every $n=0,1,\cdots$. If $f(t)$ is an L-function of order n (cf. [3], p. 17), then the function on the left of (43) is also an L-function of order n. But an L-function of order n cannot be asymptotically equal to $1/l_{n+1}{}^2(t)$; cf. [2], p. 81. Hence, $M(f)$ exists and does not exceed the order of f.

6. Proof of Lemma 2. The assumption that (3) satisfies $|\lambda|>0$ implies that (14_0) holds, so that (15_0) defines $F_1(t)$ for large t. From (11), (15_0) and (12_0),

$$(\pm 1-F_1(t))\,|f(t)|^{\frac{1}{2}}=-|f|^{-1/4}\,d^2\,|f|^{-1/4}/dt^2,$$

which can be written in the form

$$g''/4g^{5/2}-5g'^2/16g^{5/2},$$

if $g=|f|$ and the prime denotes differentiation with respect to t. Since the two terms in the last formula are L-functions and, therefore, either identically zero or of constant sign for large t, the relation (22) will follow if it is shown that

$$\int^t (g''/g^{3/2})\,dt \quad\text{and}\quad \int^t (g'^2/g^{5/2})\,dt \tag{44}$$

tend to finite limits as $t\rightarrow\infty$.

First

$$\int^t (g'/g^{5/4})\,dt=\text{const.}-4g^{-1/4}(t)$$

tends to a finite limit since $|\lambda|>0$. Since $g'/g^{5/4}$ is (improperly) integrable and monotone for large t-values, it follows that

$$g'/g^{5/4}\rightarrow 0 \text{ as } t\rightarrow\infty. \tag{45}$$

Hence,

$$g'^2/g^{5/2}=(g'/g^{5/4})^2$$

is (improperly) integrable. The second expression in (44) tends, therefore, to a finite limit.

In a manner similar to that employed in obtaining (45), it is seen that

$$g'/g^{3/2}\rightarrow 0 \text{ as } t\rightarrow\infty.$$

An integration by parts shows that the first expression in (44) equals

$$\text{const.} + g'/g^{3/2} + (3/2)\int^{t} (g'^{2}/g^{5/2})\,dt.$$

Since this function tends to a finite limit, as $t \to \infty$, the assertion (22) is proved.

The relation (21) is a consequence of (22), for the integrand in (22) is an integrable L-function and, consequently, tends to 0 as $t \to \infty$: On the other hand, the factor $|f(t)|^{\frac{1}{2}}$ stays away from 0 by virtue of the assumption $|\lambda| > 0$. This completes the proof of Lemma 2.

7. Proof of Lemma 3. The case $n = 1$ of the Lemma 3 will first be proved. Suppose that (23_1) holds, that is, that

$$F_1(t) = o(1). \tag{46}$$

It will be shown that (24_1) holds, that is, that $\lambda_0 = 0$ or, equivalently,

$$f(t) = (1 + o(1))/4t^2. \tag{47}$$

From (11), (12_0) and (15_0),

$$F_1(t) = \pm 1 + |f|^{-3/4} d^2 |f|^{-1/4}/dt^2. \tag{48}$$

Hence, (46) shows that

$$d^2h/dt^2 = -(\pm 1 + o(1))/h^3, \tag{49}$$

where $h = |f|^{-1/4}$. The assumption (46) and Lemma 2 implies that

$$h = |f|^{-1/4} \to \infty \text{ as } t \to \infty. \tag{50}$$

If (49) is multiplied by dh/dt, a quadrature gives

$$\tfrac{1}{2}(dh/dt)^2 = \tfrac{1}{2}(\pm 1 + o(1))/h^2 + \text{ const.} \tag{51}$$

Actually, the integration constant must be 0, by virtue of (14_0). For by (50),

$$\lim_{t\to\infty} \tfrac{1}{2}(dh/dt)^2 = \text{const.},$$

so that const. $\geqq 0$. However, if const. > 0, then

$$h \sim (2\text{ const.})^{\frac{1}{2}} t$$

or

$$f \sim 1/(2\text{ const.})^2 t^4. \tag{52}$$

This contradicts (14_0) in virtue of (11) and (12_0).

Hence, const. $= 0$ in (51). Therefore,

(53) $$\pm 1 = \operatorname{sgn} f(t) = 1 \text{ for large } t$$

and

$$h\, dh/dt = 1 + o(1)$$

or

$$\tfrac{1}{2}h^2 = (1 + o(1))t.$$

Finally, the definition of h and (53) show that (47) holds. This proves that (46) implies (47) when (14_0) holds.

The converse is readily verified. In fact, the asymptotic relation (47) can be differentiated (twice). If use is made of (48), the relation (46) is then deduced. This completes the proof of the Lemma 3 in the case $n = 1$.

Assume that the Lemma 3 is true for arbitrary $n \geqq 1$. It will be shown that it is true for $n + 1$, that is, that if $(14_0), (14_1), \cdots, (14_n), (23_n)$ and (24_n) hold, then

(54) $$\lim_{t\to\infty} F_{n+1}(t) = 0$$

holds if and only if

(55) $$\lambda_n = 0.$$

The assumptions (23_n) and/or (24_n) mean that

(56) $$f(t) = (1/4t^2)\Big(\sum_{j=0}^{n-1} L_j(t) + \epsilon_{n-1}(t)L_{n-1}(t)\Big),$$

where

(57) $$\epsilon_{n-1}(t) = o(1).$$

Since $L_0(t) \equiv 1$, it is seen that

$$|f(t)|^{\frac{1}{2}} = (1/2t)\Big(1 + \tfrac{1}{2}\sum_{j=1}^{n-1} L_j(t) + o(1)L_{n-1}(t)\Big).$$

Let $t_1 = t_1(t)$ be defined by

(58) $$t_1 = t_1(t) = \int^t |f(s)|^{\frac{1}{2}}\, ds,$$

where the lower limit of integration, say T, is fixed so that $f(t) > 0$ for $t \geqq T$. Then the last two formula lines and the definition (6) of $L_k(t)$ show that

(59) $$t_1(t) = \tfrac{1}{2}\log t + \text{const.} + o(1).$$

Let $t = t(t_1)$ denote the inverse function of (58). If $P_n^*(t_1)$ and $F_n^*(t_1)$ belong to

(60) $$F_0^*(t_1) = F_1(t(t_1))$$

in the same way as (12_n) and (15_n) belong to $F_0(t)$, then

(61) $$F_k^*(t_1) = F_{k+1}(t(t_1)) \text{ for } k = 0, 1, \cdots, n.$$

The assumptions (14_k) for $k = 1, \cdots, n$ imply that

$$\int^\infty P_k^*(t_1)\,dt_1 = \infty \text{ for } k = 0, \cdots, n-1,$$

since (58), (61) and (12_n) show that

$$\int^{t_1} P_k^*(t_1)\,dt_1 = \int^t P_{k+1}(t)\,dt.$$

Assume first that (54) and (23_n) and/or (24_n) hold. It will be proved that (55) holds. Since $t \to \infty$ and $t_1 \to \infty$ are equivalent by (59), it is seen from (23_n) and (54) that

(62_n) $$\lim_{t_1\to\infty} F_k^*(t_1) = 0 \text{ for } k = 1, \cdots, n.$$

Hence by the induction hypothesis

(63_n) $$\lambda_k^* = 0 \text{ for } k = 0, 1, \cdots, n-1$$

if λ_n^* belongs to (60) in the same way as (9_n) does to (11). Consequently,

(64) $$F_0^*(t_1) = (1/4t_1^2)\sum_{j=0}^{n-1} L_j(t_1) + o(1)L_{n-1}(t_1)/t_1^2.$$

But (59) and (60) show that this implies

(65) $$F_1(t) = \sum_{j=0}^{n-1} L_{j+1}(t) + o(1)L_n(t).$$

On the other hand, (56) gives

$$f^{-1/4} = 4^{1/4}t^{\frac{1}{2}}\{1 - \Lambda(t)/4 + O(1)/\log^4 t\},$$

where

$$\Lambda(t) = \sum_{j=1}^{n-1} L_j(t) + \epsilon_{n-1}(t)L_{n-1}(t).$$

Differentiation of the last asymptotic formula for $f^{-1/4}$ shows that its second derivative is

$$-4^{-3/4}t^{-3/2}\{1 - \Lambda/4 + (t\epsilon' + t^2\epsilon'' + o(1)\epsilon + o(1)t\epsilon')L + O(1)/\log^3 t\},$$

where $\epsilon = \epsilon_{n-1}$ and $L = L_{n-1}$, since the derivative of an $O(1)$ L-function is $o(1/t)$. Also, (56) gives that

$$f^{-3/4} = 4^{3/4}t^{3/2}\{1 - 3\Lambda/4 + O(1)/\log^4 t\}.$$

Hence, (48) and (53) lead to

$$F_1(t) = \Lambda - (t\epsilon' + t^2\epsilon'' + o(1)\epsilon + o(1)t\epsilon')L + O(1)/\log^3 t.$$

A comparison of this relation with (65) shows that

$$(\epsilon_{n-1} - t\epsilon_{n-1}' - t^2\epsilon_{n-1}'' + o(1)\epsilon_{n-1} + o(1)t\epsilon_{n-1}')L_{n-1} \sim L_n$$

or that

$$\epsilon_{n-1} - (t\epsilon_{n-1}' + t^2\epsilon_{n-1}'') + o(1)\epsilon_{n-1} + o(1)t\epsilon_{n-1}' \sim 1/l_n^2(t).$$

If the new independent variable $s = \log t$ is introduced, the last formula becomes

$$\epsilon_{n-1} - d^2\epsilon_{n-1}/ds^2 + o(1)\epsilon_{n-1} + o(1)ds_{n-1}/ds \sim 1/l_{n-1}^2(s).$$

Since $\epsilon_{n-1} = \epsilon_{n-1}(e^s)$ and its derivatives tend monotonously to zero, it follows that, as $s \to \infty$,

$$d\epsilon_{n-1}/ds = o(1/s) \text{ and } d^2\epsilon_{n-1}/ds^2 = o(1/s).$$

Hence,

$$\epsilon_{n-1}(e^s) \sim 1/l_{n-1}^2(s) \text{ or } \epsilon_{n-1}(t) \sim 1/l_n^2(t).$$

The equality (55) follows, therefore, from (56) and (9_n).

It remains to prove the converse, namely, that (55) and (23_n) and/or (24_n) imply (54). Since the reasoning of the preceding paragraph can be reversed, it is seen that (55) and (24_n) imply the relation (65) or, by virtue of (59), the relation (64). But (64) means that (63_n) holds, which implies (62_n) by the induction hypothesis. Consequently, (54) follows from (61).

8. Proof of Lemma 4. This lemma is a consequence of the proof of the last lemma. Thus, the assumption $N(f) > 0$ implies (47), i.e., (25_0); so that if $N \geqq 1$, it follows from (12_0) that

$$\int^t P_0(t)\,dt \sim \tfrac{1}{2}\log t,$$

i.e., that (26) holds. If $N(f) > 1$, then (25_1) follows from (65); while

(25_n) can be proved for arbitrary n by induction. Finally, (26) is a consequence of (25_n) for $0 \leqq n < N$.

9. **Proof of Lemma 5.** The lemma will first be proved for the case when $N = N(f) = 0$. Suppose first that λ_0 is finite but $\lambda_0 \neq -1$. Then (9_0) means

$$f \sim (1 + \lambda_0)/4t^2$$

or

$$| f |^{-1/4} \sim 4^{1/4} t^{\frac{1}{2}} / | 1 + \lambda_0 |^{1/4}.$$

Since this asymptotic relation can be differentiated (twice),

$$d^2 | f |^{-1/4} / dt^2 \sim -1/(4^{3/4} t^{3/2} | 1 + \lambda_0 |^{1/4}),$$

so that

$$| f |^{-3/4} d^2 | f |^{-1/4} / dt^2 \sim -1/ | 1 + \lambda_0 |.$$

Since the ± 1 in (48) is $\operatorname{sgn} f = \operatorname{sgn}(1 + \lambda_0)$, it is seen that

$$\lim F_1(t) = (1 + \lambda_0)/ | 1 + \lambda_0 | - 1/ | 1 + \lambda_0 | = \lambda_0 / | 1 + \lambda_0 |.$$

This completes the proof in the case λ_0 is finite but not -1.

If $| \lambda_0 | = \infty$, then (9_0) means

$$4t^2 | f | \to \infty \quad \text{or} \quad | f |^{-1/4} = o(t^{\frac{1}{2}}).$$

The last asymptotic relation can be differentiated and so

$$| f |^{-3/4} d^2 | f |^{-1/4} / dt^2 = o(1).$$

In view of (48), this establishes the case $N = 0$ and $| \lambda_0 | = \infty$, since $\pm 1 = \operatorname{sgn} f$ is $+1$ or -1 according as $\lambda_0 = \infty$ or $\lambda_0 = -\infty$.

Finally, if $\lambda_0 = -1$, then (9_0) means

$$(67) \qquad 4t^2 f = o(1) \quad \text{or} \quad t^{\frac{1}{2}} = o(| f |^{-1/4}).$$

But since (16) does not hold, it is seen from (12_0) and (11) that

$$| f |^{-1/4} = o(t^{\frac{1}{2}+\epsilon}), \qquad (\epsilon > 0 \text{ arbitrary}),$$

so that, for large t, the graph of

$$| f |^{-1/4} \text{ is concave downwards}$$

and

$$d\, | f |^{-1/4}/dt = o(1).$$

The last asymptotic relation in (67) can, therefore, be differentiated twice and shows that

$$| f |^{-3/4} d^2 \, | f |^{-1/4}/dt^2 \to -\infty.$$

This completes the verification of the case $N = 0$ of Lemma 5.

The proof can be completed by an induction on N. Assume the lemma is true for functions $f(t)$ for which $N(f) = N$, it will be shown that it holds for functions $f(t)$ satisfying $N(f) = N + 1$. Let $N(f) = N + 1 \geqq 1$. Introduce the notations of the proof of Lemma 3. Clearly, $N(F_0{}^*) = N(f) - 1 = N$ in virtue of the proof of Lemma 3 and the negation of (16); cf. the formulae following (61). Consequently, the induction hypothesis implies

$$\lim_{t_1 \to \infty} F_N{}^*(t_1) = \lambda_N{}^*/|\, 1 + \lambda_N{}^* \,| \neq 0. \tag{68}$$

In view of (60),

$$F_1(t) - \sum_{j=0}^{k} L_{j+1}(t) = o(L_{k+1}(t)) \text{ for } k = 1, \cdots, N - 1$$

and

$$(F_1(t) - \sum_{j=0}^{N} L_{j+1}(t))/L_{N+1}(t) \to \lambda_N{}^* \neq 0.$$

Proceeding as in the proof of Lemma 3, one verifies that

$$\epsilon_N(t)\, l_{N+1}{}^2(t) - 1 \to \lambda_N{}^*.$$

In view of (56), where $n - 1 = N$, it is seen from (9_{N+1}) that $\lambda_{N+1} = \lambda_N{}^*$. Hence, by (61) and (68),

$$\lim_{t \to \infty} F_{N+1}(t) = \lambda_{N+1}/|\, 1 + \lambda_{N+1} \,|.$$

This completes the induction and the proof of Lemma 5.

10. Remarks. The possible asymptotic formulae (4_{iii}), (17), (18), (20) for the solutions of (1) furnish quite simply the answers to such questions

as the existence or non-existence of oscillatory solutions, of unbounded solutions, of $o(1)$-solutions, of $L^2(0, \infty)$-solutions, etc.

As an illustration, consider A. Kneser's theorem [7], p. 415, stating that if $q(t)$ is continuous, then

$$y'' + q(t)y = 0 \tag{69}$$

possesses an oscillatory solution $y = y(t)$ if

$$\liminf_{t\to\infty} (4t^2q(t) - 1) > 0,$$

and does not possess an oscillatory solution if

$$\limsup_{t\to\infty} (4t^2q(t) - 1) < 0.$$

The case where

$$\liminf_{t\to\infty} (4t^2q(t) - 1) < 0 < \limsup_{t\to\infty} (4t^2q(t) - 1),$$

or

$$\lim_{t\to\infty} (4t^2q(t) - 1) = 0$$

is undecided. The results obtained above show that in this latter case, the function

$$4t^2q(t) - 1 - 1/\log^2 t$$

should be considered. If this function has a positive (negative) limit inferior (superior), then (69) does (not) possess an oscillatory solution. If the limit superior and inferior are both 0, the function

$$4t^2q(t) - 1 - 1/\log^2 t - 1/(\log\, \log t)^2 \log^2 t$$

should be considered; etc.

Furthermore, it is easy to see from (18) and the remark concerning (19) that if (1) is oscillatory, then either no solution $x = x(t) \not\equiv 0$ of (1) or every solution is of class $L^2(0, \infty)$ according as

$$\int^{\infty} f^{-\frac{1}{2}}(t)\, dt$$

is ∞ or finite; cf. [5], p. 306.

It is also easy to see that if (1) is oscillatory and $f(t) \to 0$, then every solution $x = x(t) \not\equiv 0$ of (1) is unbounded; cf. [6].

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Bôcher, "On regular singular points of linear differential equations of the second order whose coefficients are not necessarily analytic," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 40-52.

[2] G. H. Hardy, "Properties of logarithmico-exponential functions," *Proceedings of the London Mathematical Society*, ser. 2, vol. 10 (1912), pp. 54-90.

[3] ———, *Orders of Infinity*, Cambridge Tracts, no. 12, 2nd ed., (1924).

[4] P. Hartman, "On unrestricted solution fields of almost-separable differential equations," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 560-580.

[5] ——— and A. Wintner, "Criteria of non-degeneracy for the wave equation," *American Journal of Mathematics*, vol. 70 (1948), pp. 295-308.

[6] ——— and A. Wintner, "On non-conservative linear oscillators of low frequency," *American Journal of Mathematics*, vol. 70 (1948), pp. 529-539.

[7] A. Kneser, "Untersuchungen über die reellen Nullstellen der Integrale linearer Differentialgleichungen," *Mathematische Annalen*, vol. 42 (1893), pp. 409-435.

[8] A. Wintner, "Asymptotic integrations of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

[9] ———, "On the normalization of the characteristic differentials in continuous spectra," *Physical Review*, vol. 72 (1947), pp. 516-517.

[10] ———, "Asymptotic integrations of the adiabatic oscillator in its hyperbolic range," *Duke Mathematical Journal*, vol. 15 (1948), pp. 56-67.

AN APPLICATION OF SPECTRAL THEORY TO A SINGULAR CALCULUS OF VARIATIONS PROBLEM.*

By C. R. Putnam.

Introduction.

The theory of boundary value problems has been applied by Lichtenstein to problems of the calculus of variations. The object of this paper will be to show that some of his results [7] on one-dimensional problems can be carried over to certain types of singular calculus of variations problems. The chief difficulty in extending Lichtenstein's results to singular problems lies in the fact that the associated boundary value problem may be degenerate (that is, in the "Grenzkreisfall" in Weyl's terminology [9], p. 238) and that the spectrum need not consist of a sequence of eigenvalues tending to infinity, but can contain continuous and cluster spectra.

In the sequel, the following terminology will be used: A real-valued function $y = y(x)$, where $0 \leqq x < \infty$, will be said to belong to class Ω_α if

(i) $y(x)$ is continuous for $0 \leqq x < \infty$;

(ii) the half-line $0 \leqq x < \infty$ can be divided into a sequence of intervals $0 \leqq x \leqq a_1$, $a_1 \leqq x \leqq a_2, \cdots$, where $a_n \to \infty$ as $n \to \infty$, in such a way that $y(x)$ possesses a continuous derivative $y'(x)$ on each of these intervals;

(iii) $y(x)$ satisfies the boundary condition

(1) $$y(0) \cos \alpha + y'(0) \sin \alpha = 0, \qquad (0 \leqq \alpha < \pi);$$

(iv) $y(x)$ is of class (L^2), that is,

(2) $$\int_0^\infty y^2 dx < \infty.$$

In the differential equation,

(3) $$y'' + (\lambda - q) y = 0,$$

let $q = q(x)$ be a real continuous function for $0 \leqq x < \infty$ and let λ denote a real parameter. If (3) is in the "Grenzpunktfall," that is, if for some λ (hence, for all λ), the differential equation possesses at least one solution

* Received April 14, 1948.

$y = y(x)$ which fails to satisfy (2), then the set of λ-values in the spectrum of the boundary value problem belonging to (3) and the boundary condition (1) will be denoted by S_α. As to the definition of "spectrum," see [9], p. 251 or 1 below.

The main tool in this paper will be a generalization (Lemma 2 below), of the Parseval identity which plays the principle role in Lichtenstein's investigations, cf. [7], p. 166.

This paper will consist of two parts. In Part II, the Lemma will be applied to problems of locating the spectra of certain boundary value problems. For example, it will be used to show that if

$$q(x) \to c \text{ as } x \to \infty, \tag{4}$$

then $\lambda = c$ belongs to the spectrum S_α for every α. That is, if condition (4) is fulfilled, then $\lambda = c$ is a cluster point of the spectrum S_α for every α (cf. [9], p. 251).

Let $\lambda_0(X) < \lambda_1(X) < \cdots$ denote the sequence of eigenvalues belonging to the Sturm-Liouville problem determined by the differential equation (3), the boundary condition (1) at $x = 0$ and the boundary condition

$$y(X) = 0 \qquad (0 < X < \infty). \tag{5}$$

Let the function $q = q(x)$, where $0 \leqq x < \infty$, be continuous and bounded from below,

$$q(x) > \text{const. for } 0 \leqq x < \infty. \tag{6}$$

Then, the differential equation (3) is in the Grenzpunktfall and the closed set S_α is bounded from below (Weyl [9], p. 238). If λ_0 denotes the least point in the spectrum S_α, it will be shown that $\lambda_0(X) \to \lambda_0$ as $X \to \infty$. This is the analogue of the theorem of Wintner [11], p. 231, that the end-points of the spectrum of a half-bounded matrix are the limits of the end-points of the spectra of the finite section matrices. If the least point λ_0 of the spectrum S_α is isolated, the next point of the spectrum can be located in a similar manner. In particular, an analogue of the theorem of Milne [8] which deals with the case that S_α consists of isolated eigenvalues will be deduced. These results touch upon the more difficult problem of deciding whether S_α is a "limit," as $X \to \infty$, in some sense, of the spectra belonging to Sturm-Liouville problems on $0 \leqq x \leqq X$.

If the function $q = q(x)$ satisfies the unilateral restriction (6), it will be shown (2) that the integral

$$J(y) = \int_0^\infty (y'^2 + qy^2)\,dx \tag{7}$$

exists (possibly as $+\infty$) as an improper Riemann integral for every function $y(x)$ of class Ω_α.

The principal type of problem considered in Part I is to determine necessary and sufficient conditions that a function $y = y(x)$ furnish a minimum to the integral (7) with respect to a certain class of functions. The results obtained for such a problem have applications in the consideration of the second variation for more general types of singular calculus of variations problems. It will be proved that the integral (7) satisfies

$$\infty \geqq J(y) \geqq 0 \tag{8}$$

for all functions $y = y(x)$ of class Ω_0 if and only if the spectrum S_0 contains no negative values of λ. As an application of this result and its variants, there will be derived (**5**, **6** below) known inequalities, as for example, the standard inequality occurring in quantum mechanics, cf. Weyl [10], p. 272. It will also be shown that the result concerning (8) leads to necessary and sufficient conditions in the isoperimetric problem of minimizing (7) with respect to the functions $y(x)$ in Ω_0 which are normalized by

$$\int_0^\infty y^2 dx = 1. \tag{9}$$

Part I.

1. Let $\phi(x, \lambda, \alpha)$ denote the unique solution of the differential equation (3) determined by the initial condition

$$\phi(0, \lambda, \alpha) = -\sin\alpha \text{ and } \phi'(0, \lambda, \alpha) = \cos\alpha,$$

where the prime denotes partial differentiation with respect to x. If (3) is in the Grenzpunktfall, the number λ is said to be an eigenvalue of the boundary value problem determined by (3) and (1) if $\phi(x, \lambda, \alpha)$ is of class (L^2); in which case, any non-zero constant multiple of $\phi(x, \lambda, \alpha)$ is called an eigenfunction belonging to λ. There exists at most a denumerable set of eigenvalues $\lambda_1, \lambda_2, \cdots$ for any fixed α. Weyl's application [9], pp. 239-251, of the spectral resolution theory of Hellinger [6] implies that there exists a monotone non-decreasing function $\rho = \rho_\alpha(\lambda)$, where $-\infty < \lambda < \infty$, which has the properties that:

(i) $\rho_\alpha(0) = 0$

(ii) $\rho_\alpha(\lambda)$ is continuous from the left and its discontinuity points are the eigenvalues $\lambda = \lambda_j$.

(iii). Let $\rho_\alpha^*(\lambda)$ denote the continuous component of $\rho_\alpha(\lambda)$ and let

$$P_1(x,\lambda,\alpha)=\int_0^\lambda \phi(x,\mu,\alpha)\,d\rho_\alpha^*(\mu),$$

where the integral is a Riemann-Stieltjes integral. Let

$$a_j=1/\int_0^\infty \phi^2(x,\lambda_j,\alpha)\,dx,$$

so that $a_j^{\frac{1}{2}}\phi(x,\lambda_j,\alpha),\ j=1,2,\cdots$, form an ortho-normal sequence, and let $d\rho_\alpha(\lambda_j)$ denote the jump $\rho_\alpha(\lambda_j+0)-\rho_\alpha(\lambda_j)$ of $\rho_\alpha(\lambda)$ at $\lambda=\lambda_j$. If Δ denotes a closed λ-interval $\lambda_1\leqq\lambda\leqq\lambda_2$ and if $f(\lambda)$ is an arbitrary function of λ for which every point is of the first kind, then by Δf will be meant $f(\lambda_2+0)-f(\lambda_1-0)$, with the understanding that the interval Δ may reduce to a single point. Let the $d\rho_\alpha(\lambda_j),\ j=1,2,\cdots$, be chosen so that

$$\sum_j (a_j d\rho_\alpha(\lambda_j))^{\frac{1}{2}}<\infty.$$

Let $P_2(x,\lambda,\alpha)$ denote the function of class (L^2) such that $P_2(x,0,\alpha)=0$ for all x and α and such that $\Delta P_2(x,\lambda,\alpha)$ is defined by the "Fourier" series

$$\Delta P_2(x,\lambda,\alpha)=\sum_{\lambda_1<\lambda_j\leqq\lambda_2}(a_j d\rho_\alpha(\lambda_j))^{\frac{1}{2}}\phi(x,\lambda_j,\alpha).$$

In virtue of the properties of $\phi(x,\lambda,\alpha)$, cf. [9], p. 222, it follows that $\Delta P_2'$ and $\Delta P_2''$ may be obtained by term-by-term differentiations of this series. If $P(x,\lambda,\alpha)$ is defined by

$$P(x,\lambda,\alpha)=P_1(x,\lambda,\alpha)+P_2(x,\lambda,\alpha), \tag{10}$$

then, for any fixed α, $P(x,\lambda,\alpha)$ is an orthogonal differential system with basis $\rho_\alpha(\lambda)$, cf., e.g., Carleman [1], p. 17. Thus,

$$\int_0^\infty \Delta_1 P\Delta_2 P\,dx=\Delta_{12}\rho_\alpha,$$

where the integral exists as a Lebesgue integral for all λ-intervals Δ_1, Δ_2 and where Δ_{12} denotes the common part of Δ_1 and Δ_2.

(iv) For every fixed α, the orthogonal differential system $P(x,\lambda,\alpha)$ is complete; that is, if $y=y(x)$ is of class (L^2) and if $\Gamma(\lambda,\alpha)$ is defined by

$$\Delta\Gamma(\lambda,\alpha)=\int_0^\infty y(x)\Delta P(x,\lambda,\alpha)\,dx \tag{11}$$

and $\Gamma(0,\alpha)=0$, then the Parseval identity

$$\int_0^\infty y^2(x)\,dx=\int_{-\infty}^\infty (d\Gamma)^2/d\rho_\alpha(\lambda) \tag{12}$$

holds. The integral in (11) exists as a Lebesgue integral since y and ΔP are of class (L^2) and the last integral in (12) exists as a Hellinger integral [5], pp. 25-28.

(v) Finally, the function $P_1(x, \lambda, \alpha)$ of (10) satisfies the differential equation

$$L(P_1) + \int_0^\lambda \mu dP_1(x, \mu, \alpha) = 0, \tag{13}$$

that is,

$$L(P_1) + \lambda P_1 = \int_0^\lambda P_1(x, \mu, \alpha)\, d\mu, \tag{13 bis}$$

where $L(y)$ is the differential operator

$$L(y) = y'' - qy. \tag{14}$$

The continuous component $\rho_\alpha^*(\lambda)$ of $\rho_\alpha(\lambda)$ is unique up to an additive constant. The spectrum S_α of the boundary value problem belonging to (3) and (1) consists of the point-spectrum, $\lambda_1, \lambda_2, \cdots$, and their cluster points (cluster spectrum) and the continuous spectrum; the latter being defined as the set of points λ, in every neighborhood of which $\rho_\alpha^*(\lambda)$ is not constant.

For later reference, it may be pointed out that the function of x on the right of (13 bis) is of class (L^2), for every fixed α and λ.

2. If $q = q(x)$ satisfies (6), it has been previously remarked that the set S_α is bounded from below. Hence, if $y = y(x)$ is of class (L^2) and if $\Gamma(\lambda, \alpha)$ is defined by (11), then the Hellinger integral

$$\int_{-\infty}^{\infty} \lambda (d\Gamma)^2 / d\rho_\alpha(\lambda)$$

exists (with the understanding that $+\infty$ is an admissible value). Similarly, if $y(x)$ is of class Ω_α, then the integral (7) exists as an improper Riemann integral, again with the understanding that its value may be $+\infty$. This is clear from the fact that the integral of y'^2 is non-negative; the integral of qy^2 over the set of points x where $q(x) < 0$ is finite in virtue of (2) and (6); finally, the integral of qy^2 over the set of points x where $q(x) \geqq 0$ is non-negative.

Suppose that the function $y(x)$ is of class Ω_α and possesses a continuous second derivative such that $L(y)$ is of class (L^2). If $\Lambda(\lambda, \alpha)$ is the "Fourier" transform of $L(y)$,

$$\Delta\Lambda(\lambda, \alpha) = \int_0^\infty L(y) \Delta P(x, \lambda, \alpha)\, dx, \tag{15}$$

then, by the Green identity,

$$\Delta\Lambda(\lambda,\alpha) = \int_0^\infty yL(\Delta P)\,dx + [y'\Delta P - y\Delta P']_0^\infty.$$

In the sequel, the notation

$$f(x)\,\Big|_0^\infty = \lim_{x\to\infty} f(x) - f(0)$$

will be used, whenever the limit involved exists. It follows from a remark of Weyl [9], pp. 241-242, that the corresponding limit in the Green identity following (15) does exist and equals 0. The contribution of $x=0$ to the Green identity is 0, since y and ΔP satisfy the same boundary condition (1). By (10) and (13), the relation

$$\Delta\Lambda(\lambda,\alpha) + \lambda^*\Delta\Gamma(\lambda,\alpha) = 0 \tag{16}$$

between Γ and Λ, the Fourier transforms of y and $L(y)$, holds, where λ^* is some point of the λ-interval Δ. In virtue of (11), (15) and (16), an application of the polarized form of the Parseval identity gives

$$-\int_0^\infty yL(y)\,dx = \int_{-\infty}^\infty \lambda(d\Gamma)^2/d\rho_\alpha(\lambda). \tag{17}$$

An integration by parts and (14) show that this last relation is equivalent to

$$\int_0^\infty (y'^2+qy^2)\,dx = y(x)y'(x)\,\Big|_0^\infty + \int_{-\infty}^\infty \lambda(d\Gamma)^2/d\rho_\alpha(\lambda), \tag{18}$$

provided the integral on the left exists as an improper Riemann integral.

The following generalization of the Parseval identity of Lichtenstein will be proved:

LEMMA. *If $q=q(x)$ is continuous on $0\leqq x<\infty$ and satisfies the condition* (6), *and $y=y(x)$ belongs to an arbitrary Ω_α, then*

$$\infty \geqq J(y) \geqq -y(0)y'(0) + \int_{-\infty}^\infty \lambda(d\Gamma)^2/d\rho_\alpha(\lambda). \tag{19}$$

The last equality sign in (19) is surely valid in case y belongs to Ω_α, $J(y)$ exists as a finite integral and y possesses a continuous second derivative such that $L(y)$ is of class (L^2). This is clear from relations (17) and (18); for these restrictions imply that both y and y' are of class (L^2) and, consequently, that

$$\lim_{x\to\infty} y(x)y'(x)$$

exists and is 0. If y merely belongs to Ω_α, it remains undecided whether

or not the last equality sign is always valid. This circumstance, however, is of no consequence in the application of (19) to be made below.

3. Proof of the Lemma. It is clear from the discussion at the beginning of 2 that

(20) $$\int_0^\infty y'^2 dx < \infty \text{ and } \int_0^\infty |q| y^2 dx < \infty$$

can be assumed.

Since $y = y(x)$ is of class (L^2), there exists a sequence of x-values $x_1 < x_2 < \cdots$ such that

(21) $$|y(x_n)| \to 0 \text{ and } 1 < x_n \to \infty, \qquad (n \to \infty).$$

Let $\epsilon_1 > \epsilon > \cdots$ be a sequence of positive numbers satisfying

(22) $$1 > \epsilon_n \to 0, \qquad (n \to \infty).$$

Let $v_n = v_n(x)$ be any function possessing a continuous first derivative on the interval $0 \leqq x \leqq x_n$ and satisfying the relations

(23) $$v_n(0) = y'(0), \; v_n(x_n) = 0$$

and

(24) $$\int_0^{x_n} (y' - v_n)^2 dx < \epsilon_n / n^2 M_n,$$

where M_n is defined by

(25) $$M_n = x_n^2 \max_{0 \leqq x \leqq x_n} (1, |q(x)|).$$

Let $y_n = y_n(x)$ be defined by

(26) $$y_n(x) = \int_0^x v_n(t)\,dt + y(0), \text{ where } 0 \leqq x \leqq x_n.$$

It follows that $y_n(x)$ has a continuous second derivative; that

(27) $$y_n(0) = y(0);$$

and that

(28) $$y_n(x_n) \to 0, \qquad (n \to \infty).$$

The last relation is a consequence of the fact that

(29) $$y(x) - y_n(x) = \int_0^x (y'(t) - v_n(t))\,dt$$

and, therefore, for $0 \leqq x \leqq x_n$,

(30) $$(y(x) - y_n(x))^2 \leqq x_n \int_0^{x_n} (y'(t) - v_n(t))^2 dt < \epsilon_n x_n / n^2 M_n,$$

by (24). The limit relation (28) now follows from (21), (25) and (22); in fact;

(31) $$|y_n(x_n)| < |y(x_n)| + 1/n.$$

Since (30) is valid for $0 \leqq x \leqq x_n$, the definition (25) of M_n shows that

(32) $$\int_0^{x_n} (y - y_n)^2 dx < \epsilon_n$$

and

(33) $$\int_0^{x_n} |q| (y - y_n)^2 dx < \epsilon_n.$$

Hence, (22), (24), (26) and (33) imply

(34) $$\left| \int_0^{x_n} (y_n'^2 + qy_n^2) dx - \int_0^{x_n} (y'^2 + qy^2) dx \right| \to 0, \qquad (n \to \infty).$$

The definition of the function $y_n(x)$ for $0 \leqq x \leqq x_n$, given in (26), will now be extended over the half-axis $0 \leqq x < \infty$ in such a way that $y_n(x)$ possesses a continuous second derivative and satisfies

(35) $$y_n(x) \equiv 0 \text{ for } x_n + 1 \leqq x < \infty,$$

and

(36) $$\int_0^{\infty} (y - y_n)^2 dx \to 0, \qquad (n \to \infty);$$

finally,

(37) $$\int_0^{\infty} (y_n'^2 + qy_n^2) dx \to \int_0^{\infty} (y'^2 + qy^2) dx, \qquad (n \to \infty).$$

In virtue of (32) and (34), it is sufficient to construct $y_n(x)$ with the desired smoothness and with the properties that

$$\int_{x_n}^{x_n+1} y_n^2 dx \to 0 \text{ and } \int_{x_n}^{x_n+1} |q| y_n^2 dx \to 0, \qquad (n \to \infty),$$

and

$$\int_{x_n}^{x_n+1} y_n'^2 dx \to 0, \qquad (n \to \infty).$$

If θ_n is defined by

(38) $$\theta_n = 2 \max (|y(x_n)|, 1/n),$$

then

(39) $$0 < \theta_n \to 0, \qquad (n \to \infty),$$

by (21). Let $s(x)$ denote the linear function defined on the interval $x_n \leqq x \leqq x_n + 1$, the graph of which is the line segment joining the points

with coordinates (x_n, θ_n) and $(x_n+1, 0)$. Define $y^* = y^*(x)$ on $x_n \leqq x \leqq x_n + 1$ by placing

$$y^*(x) = \begin{cases} \min\,(y(x), s(x)) \text{ if } y(x) \geqq 0, \\ \max\,(y(x), -s(x)) \text{ if } y(x) \leqq 0. \end{cases} \tag{40}$$

Since $y(x)$ and $s(x)$ are absolutely continuous, y^* is also and, therefore, $y^{*\prime}$ exists almost everywhere. Moreover,

$$\int_{x_n}^{x_n+1} y^{*2}dx \leqq \int_{x_n}^{x_n+1} y^2 dx, \tag{41}$$

and

$$\int_{x_n}^{x_n+1} | q | y^{*2}dx \leqq \int_{x_n}^{x_n+1} | q | y^2 dx, \tag{42}$$

finally,

$$\int_{x_n}^{x_n+1} y^{*\prime 2}dx \leqq \int_{x_n}^{x_n+1} s^{\prime 2}dx + \int_{x_n}^{x_n+1} y^{\prime 2}dx.$$

Since the first integral on the right of the last inquality tends to 0 by (39) and the definition of $s(x)$. and since the second tends to 0 by virtue of y' being of class (L^2), it follows that

$$\int_{x_n}^{x_n+1} y^{*\prime 2}dx \to 0, \qquad (n \to \infty). \tag{43}$$

Let n be fixed and let $\delta = \delta_n > 0$ be a positive number to be fixed later. Let $x_n = \xi_0 < \xi_1 < \cdots < \xi_k = x_n + 1$ be a subdivision of the interval $x_n \leqq x \leqq x_n + 1$. In view of the factor 2 in the definition, (38). of θ_n, the number ξ_1 can be chosen so near x_n that

$$| y(x) | < | s(x) | \text{ for } x_n \leqq x \leqq \xi_1. \tag{44}$$

Also, if ξ_1 is sufficiently near x_n, then

$$\int_{x_n}^{\xi_1} y^{\prime 2}dx < \delta \tag{45_1}$$

and

$$\int_{x_n}^{\xi_1} \theta_n^2 dx < \delta, \qquad \int_{x_n}^{\xi_1} | q | \theta_n^2 dx < \delta. \tag{45_2}$$

Let $p = p(x)$ be the continuous function whose graph consists of the linear segments joining the points $(x_n, y_n(x_n))$, $(\xi_1, y^*(\xi_1))$, $(\xi_2, y^*(\xi_2)), \cdots$ $(\xi_k, y^*(\xi_k))$. The inequality (31) and the relation (38) imply

$$| p(x) | < \theta_n \text{ for } x_n \leqq x \leqq x_n + 1. \tag{46}$$

The definition of $p(x)$ shows that

$$\int_{x_n}^{\xi_1} p'^2 dx = (y^*(\xi_1) - y_n(x_n))^2/(\xi_1 - x_n). \tag{47}$$

From (44), $y^*(\xi_1) = y(\xi_1)$; consequently, (45_1) and the Schwarz inequality imply that

$$(y^*(\xi_1) - y(x_n))^2/(\xi_1 - x_n) < \delta.$$

Since the definitions of $y^*(x)$ and of ξ_1 do not depend on the number ϵ_n, it is clear from this last inequality and (47) that if the number ϵ_n (which so far has been arbitrary) is sufficiently small, then

$$\int_{x_n}^{\xi_1} p'^2 dx < \delta. \tag{48}$$

Clearly, (46) and (45_2) imply

$$\int_{x_n}^{\xi_1} p^2 dx < \delta \text{ and } \int_{x_n}^{\xi_1} |q|\, p^2 dx < \delta. \tag{49}$$

If the numbers k and $\xi_2, \cdots, \xi_{k-1}$ are suitably chosen,

$$|p(x) - y^*(x)| < \delta \text{ for } \xi_1 \leqq x \leqq x_n + 1,$$

since $y^*(x)$ is continuous. Hence,

$$\int_{\xi_1}^{x_n+1} p^2 dx \leqq 2 \int_{\xi_1}^{x_n+1} y^{*2} dx + 2\delta^2 \tag{50}$$

and

$$\int_{\xi_1}^{x_n+1} |q|\, p^2 dx \leqq 2 \int_{\xi_1}^{x_n+1} |q|\, y^{*2} dx + 2\delta^2 \max_{0 \leqq x \leqq x_n+1} |q(x)|. \tag{51}$$

On each of the intervals $\xi_j \leqq x \leqq \xi_{j+1}$, where $j = 1, \cdots, k-1$,

$$\int_{\xi_j}^{\xi_{j+1}} p'^2 dx = (y^*(\xi_{j+1}) - y^*(\xi_j))^2/(\xi_{j+1} - \xi_j) = (\int_{\xi_j}^{\xi_{j+1}} y^{*\prime} dx)^2/(\xi_{j+1} - \xi_j);$$

hence, an application of Schwarz's inequality on each of these intervals shows that

$$\int_{\xi_1}^{x_n+1} p'^2 dx \leqq \int_{\xi_1}^{x_n+1} y^{*\prime 2} dx. \tag{52}$$

Let $\delta = \delta_n$ be chosen so that

$$\delta_n{}^2 \max_{0 \leqq x \leqq x_n+1} (1, |q(x)|) \to 0, \qquad (n \to \infty).$$

Then

$$\int_{x_n}^{x_n+1} p'^2 dx \to 0, \qquad (n \to \infty),$$

in virtue of (52), (43) and (48). Similarly,

$$\int_{x_n}^{x_n+1} p^2 dx \to 0 \text{ and } \int_{x_n}^{x_n+1} | q | p^2 dx \to 0, \qquad (n \to \infty),$$

in virtue of (49), (50) and (51), in conjunction with (41) and (42).

The value of $p(x_n+1)$ is $y^*(x_n+1)=0$, by the definition of $p(x)$ and by (40). If the function $y_n(x)$ is defined to be $p(x)$ on $x_n \leqq x \leqq x_n+1$, then $y_n(x)$ is a continuous function for $0 \leqq x < \infty$ and satisfies the limit relations (36) and (37). Clearly, the corners of this function at $x=\xi_0$, $\xi_1, \cdots, \xi_k$ can be smoothed out without violating (35), (36) and (37). In what follows $y_n(x)$ denotes this smoothed-out function.

It follows from (35) that $y_n'(x) \equiv 0$ for $x_n+1 \leqq x < \infty$. Therefore, the boundary conditions imposed on $y_n(x)$ at $x=0$ imply that

(53) $$[y_n \Delta P' - y_n' \Delta P]_0^\infty = 0, \text{ for all } n \text{ and } \Delta,$$

and

(54) $$\int_0^\infty (y_n'^2 + q y_n^2)\, dx = -y_n(0) y_n'(0) + \int_{-\infty}^\infty \lambda (d\Gamma_n)^2 / d\rho_\alpha(\lambda),$$

where

(55) $$\Delta\Gamma_n(\lambda, \alpha) = \int_0^\infty y_n(x) \Delta P(x, \lambda, \alpha)\, dx.$$

From (37), (54) and the fact that $y_n(0) y_n'(0) = y(0) y'(0)$, it follows that

(56) $$\int_0^\infty (y'^2 + q y^2)\, dx = -y(0) y'(0) + \lim_{n\to\infty} \int_{-\infty}^\infty \lambda (d\Gamma_n)^2 / d\rho_\alpha(\lambda).$$

If $\Delta\Gamma(\lambda, \alpha)$ is defined by (15) it is seen from (36) and (55) that

$$\int_{-\infty}^\infty (d\Gamma_n - d\Gamma)^2 / d\rho_\alpha(\lambda) \to 0, \qquad (n \to \infty),$$

and, therefore,

$$\lim_{n\to\infty} \int_\gamma^\delta \lambda (d\Gamma_n)^2 / d\rho_\alpha(\lambda) = \int_\gamma^\delta \lambda (d\Gamma)^2 / d\rho_\alpha(\lambda)$$

whenever $-\infty < \gamma < \delta < \infty$. Since S_α is bounded from below, an application of Fatou's lemma to the second term on the right of the equality (56) yields the desired relation (19).

This completes the proof of the Lemma.

For later use, a corollary will be deduced.

COROLLARY. *If $q = q(x)$ is continuous on $0 \leqq x < \infty$ and satisfies the condition* (6), *and $y = y(x)$ belongs to an arbitrary Ω_β, then*

$$(57) \qquad \infty \geqq J(y) \geqq y^2(0) \cot \alpha + \int_{-\infty}^{\infty} \lambda (d\Gamma)^2 / d\rho_\alpha(\lambda), \qquad \alpha \neq 0,$$

where $\Delta\Gamma(\lambda, \alpha)$ is defined by (15).

Proof of Corollary. From the proof of the Lemma, it is clear that $y_n'(0)$ may be chosen arbitrarily, and, in particular, it may be arranged that $y_n'(0) = - y_n(0) \cot \alpha$ since $\alpha \neq 0$. Consequently, relation (53) is satisfied and, since $y_n(0) = y(0)$, the relation (57) follows from (54).

Under the same conditions as specified in the Corollary, an interesting equality is furnished by (57) for $\alpha = \frac{1}{2}\pi$, namely,

$$(58) \qquad \infty \geqq J(y) \geqq \int_{-\infty}^{\infty} \lambda (d\Gamma)^2 / d\rho_{\frac{1}{2}\pi}(\lambda).$$

Remark. Let $q(x)$ be an arbitrary continuous function on $0 \leqq x < \infty$, not necessarily subject to the restriction (6), but such that

$$\int_0^\infty |q| y^2 dx < \infty$$

and such that the differential equation (3) is in the Grenzpunktfall. Since the assumption (6) was used in the proof of the Lemma only in passing from (56) to (19), it is clear that (56) remains valid; also an equality which bears the same relation to (57) as (56) does to (19) is valid.

4. In this section, there will be proved

THEOREM (I). *If $q = q(x)$ is continuous on $0 \leqq x < \infty$ and satisfies the condition* (6), *then the integral* (7) *satisfies* (8) *for all functions $y = y(x)$ of class Ω_0 if and only if the spectrum S_0 contains no negative values of λ.*

Proof of Theorem (I). The "if" part of the theorem follows by applying the Lemma of **2**. Since $y(0)y'(0) = 0$ and the spectrum S_0 is non-negative, relation (8) results immediately from (19).

The "only if" statement follows from an application of a theorem of Wintner [12]. Suppose that there exists a negative value of λ in S_0. If there is such an eigenvalue, let y be the corresponding normalized eigenfunction. Otherwise, let $y = P_1(x, \lambda, 0)$ be chosen in such a way that

$$(59) \qquad -\int_0^\infty y L(y) dx = \int_\lambda^0 \mu d\rho_0^*(\mu) < 0.$$

In either case, the function y is of class (L^2) and satisfies a differential equation, the right member of which belongs to class (L^2), cf. equations (3) and (13 bis) and the remark at the end of **1**. In virtue of (6), the coefficient $\lambda - q$ of y in these two differential equations satisfies the inequality

$$\lambda - q < \text{const. for } 0 \leqq x < \infty.$$

It follows from the proof of the theorem [12], p. 6 (cf. also [3]), that

$$y(x)y'(x) \to 0 \text{ as } x \to \infty.$$

Since $L(y)$ is of class (L^2) and $y(0)$ is 0, it is clear (cf. (17) and (18)) that $J(y) < 0$, and the proof of Theorem (I) is complete.

A variation of the preceding theorem is contained in

THEOREM (I bis). *If $q = q(x)$ is continuous on $0 \leqq x < \infty$ and satisfies the condition (6), then the integral (7) satisfies (8) for all functions $y = y(x)$ of arbitrary Ω_a, if and only if the spectrum $S_{\frac{1}{2}\pi}$ contains no negative values of λ.*

The proof follows along the same lines as that of Theorem (I) in virtue of (58) and the fact that $P = P(x, \lambda, \frac{1}{2}\pi)$ satisfies

$$PP' \Big|_0^\infty = 0$$

for all λ.

It is clear from the two theorems above that if $S_{\frac{1}{2}\pi}$ contains no negative values of λ, then neither does S_0.

5. The results of the two preceding sections will be illustrated by the treatment of two examples. For the first of these, let

$$q(x) = (1 + ax^2)/b; \; a < 0, \; b < 0 \text{ and } ab \geqq 1,$$

so that

$$J(y) = \int_0^\infty (y'^2 + (1 + ax^2)y^2/b)\,dx, \tag{60}$$

where y belongs to an arbitrary Ω_a. Since $a/b > 0$, the function $q(x) \to \infty$ as $x \to \infty$ and the spectrum of $\rho_a(\lambda)$ consists (Weyl [9], p. 252) of an infinite sequence of eigenvalues tending to $+\infty$. The corresponding differential equation in the present case is

$$y'' + (\lambda - 1/b - ax^2/b)y = 0. \tag{61}$$

It will be shown that

$$\lambda_0 = 1/b + (a/b)^{\frac{1}{2}}$$

is a non-negative eigenvalue and the least point of $S_{\frac{1}{2}\pi}$. From $ab \geqq 1$ it follows that $\lambda_0 \geqq 0$. If $\lambda = \lambda_0$, the differential equation (61) has the solution

$$(62)\qquad y(x) = e^{-kx^2}, \text{ where } k = (a/4b)^{\frac{1}{2}} > 0.$$

This solution is of class (L^2) and satisfies the boundary condition (1) for $\alpha = \frac{1}{2}\pi$. Finally, since the solution (62) has no zeros, λ_0 is the least point of $S_{\frac{1}{2}\pi}$ (Weyl [9], p, 252). Theorem (I bis) implies that the integral $J(y)$ defined by (60) satisfies

$$(63)\qquad \infty \geqq J(y) \geqq 0,$$

for all functions y of arbitrary Ω_a.

If $-a = c$ and $-b = d$, relations (60) and (63) yield

$$(64)\qquad \int_0^\infty y^2 dx \leqq c \int_0^\infty x^2y^2 dx + d \int_0^\infty y'^2 dx,\ c > 0,\ d > 0 \text{ and } cd \geqq 1.$$

provided that the last two integrals are finite. If

$$-4cd \int_0^\infty x^2y^2 dx \int_0^\infty y'^2 dx$$

is added to both sides of the inequality obtained by squaring (64), then

$$(65)\qquad \left(\int_0^\infty y^2 dx\right)^2 - 4cd \int_0^\infty x^2y^2 dx \int_0^\infty y'^2 dx \leqq \left(\int_0^\infty (cx^2y^2 - dy'^2)\, dx\right)^2.$$

Finally, if c and d are chosen so that $cd = 1$ and that the term on the right of the inequality (65) is zero, the inequality

$$(66)\qquad \left(\int_0^\infty y^2 dx\right)^2 \leqq 4 \int_0^\infty x^2y^2 dx \int_0^\infty y'^2 dx$$

follows. It is clear that under similar assumptions, a relation occurring in quantum mechanics and corresponding to (66), but where the half-axis $0 \leqq x < \infty$ is replaced by the whole axis $-\infty < x < \infty$, is valid.

In addition, it can be remarked that when $cd\ (= ab) = 1$, then $\lambda_0 = 0$. The second inequality of (63) and the inequality of (66) then become equalities if and only if y is a constant multiple of (62).

Cf. the treatment given in Weyl [10], p. 272 and [2], pp. 165-169, 195.

6. The second example to be considered is furnished by the integral

$$(67)\qquad J(y) = \int_0^\infty (y'^2 - y^2/4(x+a)^2)\, dx,\ a > 0,$$

where y is of class Ω_0 In this case, $q(x) = -1/4(x+a)^2 \to 0$ as $x \to \infty$. By a theorem of Weyl [9], p. 252, there are no points of the spectrum S_0 in the domain $\lambda < 0$ if a non-trivial solution of the differential equation

$$(68) \qquad y'' + y/4(x+a)^2 = 0,$$

satisfying the boundary condition (1) for $\alpha = 0$, does not vanish for $0 < x < \infty$. Since

$$y(x) = (x+a)^{\frac{1}{2}}(\log(x+a) - \log a)$$

is a solution of (68) satisfying $y(0) = 0$ and since $y(x)$ has no zero on $0 < x < \infty$, the spectrum of $\rho_0(\lambda)$ is non-negative (for all $a > 0$). By Theorem (I), the integral (67) satisfies

$$(69) \qquad \infty \geqq J(y) \geqq 0,$$

for all y in Ω_0.

The restriction to consideration of functions belonging to class Ω_0 in the inequality (69) is essential. For, if $a = 1$, the differential equation (68) has a solution

$$(70) \qquad y(x) = (x+1)^{\frac{1}{2}}(-\tfrac{1}{2}\log(x+1) + 1),$$

satisfying the boundary condition (1) for $\alpha = \frac{1}{2}\pi$. Since the function (70) has a zero on $0 < x < \infty$, there exists a negative eigenvalue in the spectrum of $\rho_{\frac{1}{2}\pi}(\lambda)$ and by Theorem (I bis) the integral (67) is negative for some function y; namely, for an eigenfunction belonging to the negative eigenvalue.

If a tends to 0, it is easily shown that (69) implies

$$(71) \qquad \infty \geqq \int_0^\infty (y'^2 - y^2/4x^2)\,dx \geqq 0$$

for all y in Ω_0 for which

$$\int_0^\infty y^2/4x^2 dx < \infty.$$

Compare the present treatment of (71) with that given in [2], pp. 175-182. There, it is not assumed that y is of class (L^2) as above, but this restriction is not essential in the above considerations.

7. In this section, there will be considered the problem of minimizing the integral (7), where the function q satisfies (6) and y is of class Ω_0 and satisfies the isoperimetric condition (9).

THEOREM (II). *Let $q = q(x)$, where $0 \leqq x < \infty$, be continuous and*

satisfy (6). *There exists a function furnishing a minimum for the integral* (7), *with respect to the class of functions* y *belonging to* Ω_0 *and satisfying the isoperimetric condition* (9), *if and only if the least point* λ_0 *of the spectrum* S_0 *is an eigenvalue; in which case the corresponding normalized eigenfunction furnishes the integral* (7) *the minimum* λ_0. *If* λ_0 *is not an eigenvalue, the value* λ_0 *is not attained but is the greatest lower bound of the values of* (7) *for the class of functions specified.*

Proof of Theorem (II). Since λ_0 is the least point of S_0 it follows from (19) that

$$J(y) = \int_0^\infty (y'^2 + qy^2)\,dx \geqq \lambda_0,$$

for all functions y satisfying (9) and the boundary condition (1) for $\alpha = 0$. If λ_0 is an eigenvalue with the normalized eigenfunction y, it is seen that

$$J(y) = \lambda_0$$

and that λ_0 is, therefore, the minimum of the integral (7).

Conversely, if $J(y)$ attains a minimum value μ, then $\mu = \lambda_0$ and λ_0 is an eigenvalue. For if $\mu \neq \lambda_0$ it is clear that $\mu > \lambda_0$ and that λ_0 is not an eigenvalue. Consequently, λ_0 is a cluster point of the spectrum S_0. By choosing

$$z = z(x, \lambda) = (P(x, \lambda, 0) - P(x, \lambda_0, 0))/(\rho_0(\lambda) - \rho_0(\lambda_0))^{\frac{1}{2}}, \qquad \lambda > \lambda_0,$$

it is seen that z satisfies (9) and $J(z) \to \lambda_0$ as $\lambda \to \lambda_0$. This contradicts the assumption that μ is the minimum of (7). Hence, the minimum value of $J(y)$ is λ_0, the least point of the spectrum S_0. If the normalized function y of Ω_0 furnishes the minimum of $J(y)$, then

$$J(y) = \int_{-\infty}^\infty \lambda (d\Gamma)^2/d\rho_0(\lambda) = \lambda_0.$$

Hence, it follows from (9) and (12) that

$$\Delta\Gamma(\lambda, 0) = \int_0^\infty y(x)\Delta P(x, \lambda, 0)\,dx = 0$$

for all λ-intervals Δ not containing $\lambda = \lambda_0$. This implies that $\lambda = \lambda_0$ is an eigenvalue with the normalized eigenfunction y.

The proof of the last statement of Theorem (II) is clear from the above discussion and the proof is complete.

That the restriction (6) is essential for the validity of Theorem (II)

follows from an example of Hartman [4]. This example shows that if q does not satisfy (6), the equation (3) with $\lambda = 0$ can possess a solution y of class Ω_0 which does not vanish for $0 < x < \infty$. Consequently, the Theorem of [4] implies that (3) is in the Grenzpunktfall and that $\lambda = 0$ is the least point of the spectrum S_0. However, even though $\lambda = 0$ is an eigenvalue, it is easily verified that the integral $J(y)$ does not exist as an improper integral, where y is the normalized eigenfunction.

The last theorem of this section is

THEOREM (III). *Let $q = q(x)$, where $0 \leqq x < \infty$, be continuous and satisfy* (6). *Suppose that the least point $\lambda = \lambda_0$ of the spectrum S_0 is an eigenvalue and that $y = y_0$ denotes the corresponding eigenfunction. There exists a function furnishing a minimum for the integral* (7), *with respect to the class of functions y belonging to Ω_0 and satisfying the conditions* (9) *and*

$$\int_0^\infty y y_0 dx = 0,$$

if and only if there exists an eigenvalue $\lambda = \lambda_1 > \lambda_0$ such that no point λ of S_0 satisfies $\lambda_0 < \lambda < \lambda_1$; in which case, the normalized eigenfunction belonging to λ_1 furnishes (7) *the minimum λ_1.*

The proof of this theorem can be omitted, as it is analogous to that of the preceding theorem.

Part II.

Part II will be devoted to the derivation of information concerning the location of points of the spectral sets S_α.

8. As a result of the Lemma of **2** and the Corollary of **3**, there will first be proved

THEOREM (1). *If the spectrum S_α for a fixed α in the range $0 < \alpha \leqq \frac{1}{2}\pi$ contains no negative values of λ, then the spectrum S_β, for $\beta = 0$ and $\frac{1}{2}\pi \leqq \beta < \pi$, contains no negative values of λ.*

Proof of Theorem (1). The $\cot \alpha \geqq 0$ since $0 < \alpha \leqq \frac{1}{2}\pi$. In virtue of the hypothesis on S_α, it follows from (57) that the integral (7) satisfies (8) for all functions y of an arbitrary Ω_γ. Suppose, if possible, that $\lambda < 0$ belongs to the spectrum S_β for some β satisfying $\beta = 0$ or $\frac{1}{2}\pi \leqq \beta < \pi$. It is possible to exhibit a function y of class Ω_β (cf. the remark following the statement of the Lemma of **2**) such that the last inequality in (19) becomes

an equality while the expression appearing on the right of the equality is negative. This function y thus violates (8) and Theorem (1) is proved.

9. In this section, it will be shown, as mentioned in the introduction, that certain points in the spectrum S_a can be regarded as the limit of eigenvalues of suitable Sturm-Liouville problems. In this direction, there will be proved

THEOREM (2). *Let $q = q(x)$, where $0 \leqq x < \infty$, be continuous and satisfy* (6). *If $\lambda_0(X)$ denotes the least eigenvalue belonging to the Sturm-Liouville problem (on the finite interval $0 \leqq x \leqq X$) determined by the differential equation* (3) *and the boundary conditions* (1) *and* (5), *and λ_0 denotes the least point of the spectrum S_a of the boundary value problem determined by* (3) *and* (1), *then*

$$\lambda_0(X) \to \lambda_0, \qquad (X \to \infty). \tag{72}$$

Proof of Theorem (2). Let y be any function possessing a continuous second derivative on $0 \leqq x \leqq X$, satisfying the boundary conditions (1) and (5) and the relation

$$\int_0^X y^2 dx = 1. \tag{73}$$

Then, the completeness theorem for Sturm-Liouville problems implies that

$$-\int_0^X yL(y)\,dx = \sum_j \lambda_j c_j^2 \geqq \lambda_0(X), \tag{74}$$

where $L(y)$ is defined by (14); $\lambda_j = \lambda_j(X)$, $j = 0, 1, 2, \cdots$, are the eigenvalues, with the respective normalized eigenfunctions $\phi_j = \phi_j(x, X)$, $j = 0, 1, 2, \cdots$, of the Sturm-Liouville problem on $0 \leqq x \leqq X$ determined by (3), (1) and (5); the c_j are defined by

$$c_j = \int_0^X y(x)\phi_j(x, X)\,dx, \qquad j = 0, 1, 2, \cdots.$$

If $y = \phi_0(x, X)$ is the normalized eigenfunction belonging to $\lambda_0(X)$, the inequality in (74) becomes an equality. Let $X_1 < X_2$ and define

$$y(x) = \begin{cases} \phi_0(x, X_1) \text{ if } 0 \leqq x \leqq X_1, \\ 0 \text{ if } X_1 \leqq x \leqq X_2. \end{cases} \tag{75}$$

Then, for the function y defined by (75) on the interval $0 \leqq x \leqq X_2$, it is seen that (73), for $X = X_2$, holds and that

$$-\int_0^{X_2} yL(y)\,dx = -\int_0^{X_1} yL(y)\,dx = \lambda_0(X_1).$$

It is clearly of no consequence that the continuous function y defined by (75) does not possess a derivative at $x = X$. Since, by (74),

$$-\int_0^{X_2} yL(y)\,dx \geqq \lambda_0(X_2),$$

it follows that $\lambda_0(X_1) \geqq \lambda_0(X_2)$, that is, $\lambda_0(X)$ is a monotone, non-increasing function of X on $0 < X < \infty$. Hence,

$$\lim_{X\to\infty} \lambda_0(X) = \lambda_0(\infty)$$

exists and is not less than λ_0 by relations (17) and (12). It will be proved that $\lambda_0(\infty) = \lambda_0$, that is, that relation (72) holds. Otherwise, $\lambda_0 < \lambda_0(\infty)$ and there exists a function y (cf. proof of Theorem (I), **4**) satisfying (20) and (9) such that

$$(76) \qquad -\int_0^{\infty} yL(y)\,dx = \mu, \text{ where } \lambda_0 \leqq \mu < \lambda_0(\infty).$$

It follows from the proof in **3** of the Lemma that there exist functions $y_n = y_n(x)$, where $n = 1, 2, \cdots$, possessing continuous second derivatives, such that $y_n(0)y_n'(0) = y(0)y'(0)$, $y_n(x) \equiv 0$ if $x \geqq R_n$ ($R_n = x_n + 1$ of **3**) and (36) and (37) hold. Consequently,

$$(77) \quad \lambda_0(R_n)\int_0^{\infty} y_n^2\,dx \leqq -\int_0^{\infty} y_nL(y_n)\,dx \to -\int_0^{\infty} yL(y)\,dx, \quad (n\to\infty),$$

which, in virtue of the inequality $\lambda_0(\infty) \leqq \lambda_0(R_n)$ for all n, contradicts (76) if n is sufficiently large.

This completes the proof of Theorem (2).

THEOREM (3). *Let $q = q(x)$, where $0 \leqq x < \infty$, be continuous and satisfy (6). If the least point λ_0 of the spectrum S_α is isolated, λ_1 denotes the next larger point, and similarly $\lambda_1(X)$ is the second point of the spectrum of the Sturm-Liouville problem on $0 \leqq x \leqq X$ determined by* (3), (1) *and* (5), *then*

$$(78) \qquad \lambda_1(X) \to \lambda_1, \qquad (X\to\infty).$$

Proof of Theorem (3). Let $X_1 < X_2$ and choose constants a_0 and a_1 so that the function y defined by

$$y = \begin{cases} a_0\phi_0(x, X_1) + a_1\phi_1(x, X_1) & \text{if } 0 \leqq x \leqq X_1, \\ 0 \text{ if } X_1 \leqq x \leqq X_2 \end{cases}$$

satisfies (73) for $X = X_2$ and

$$\int_0^{X_2} y\phi_0(x, X_2)\,dx = 0. \tag{79}$$

It follows that

$$\lambda_1(X_2) \leqq -\int_0^{X_2} yL(y)\,dx = -\int_0^{X_1} yL(y)\,dx,$$

while the last integral equals

$$a_0^2\lambda_0(X_1) + a_1^2\lambda_1(X_1) \leqq \lambda_1(X_1),$$

since $a_0^2 + a_1^2 = 1$. Therefore, $\lambda_1(X)$ is a monotone, non-increasing function on $0 < X < \infty$. It is clear that

$$\lim_{X\to\infty} \lambda_1(X) = \lambda_1(\infty)$$

exists and is not less than λ_1. Suppose, if possible, $\lambda_1 < \lambda_1(\infty)$. It is convenient to suppose that λ_1 is an eigenvalue of S_α with the normalized eigenfunction $\phi_1(x)$ (otherwise, ϕ_1 will be replaced by

$$(P(x,\lambda,\alpha) - P(x,\lambda_1,\alpha))/(\rho_\alpha(\lambda) - \rho_\alpha(\lambda_1))^{\frac{1}{2}},$$

where $\lambda_1 < \lambda < \lambda_1(\infty)$). Let $\phi_0(x)$ denote the normalized eigenfunction belonging to λ_0. From the proof of the Lemma of **3**, it follows that there exist two sequences $R_0^1, R_0^2, \cdots$ and $R_1^1, R_1^2, \cdots$ and corresponding sequences of functions $\psi_0^1, \psi_0^2, \cdots$ and $\psi_1^1, \psi_1^2, \cdots$, possessing continuous second derivatives and satisfying, for $j = 0$ and 1,

$$\psi_j^n(0) = \phi_j(0), \qquad \psi_j^{n\prime}(0) = \phi_j'(0), \qquad n = 1, 2, \cdots, \tag{80}$$

and

$$\psi_j^n \equiv 0 \text{ if } x \geqq R_j^n, \qquad n = 1, 2, \cdots,$$

and, as $n \to \infty$,

$$\begin{gathered}\int_0^\infty (\psi_j^n - \phi_j)^2 dx \to 0, \quad \int_0^\infty |\,q\,|(\psi_j^n - \phi_j)^2 dx \to 0,\\ \int_0^\infty (\psi_j^{n\prime} - \phi_j')^2 dx \to 0,\end{gathered} \tag{81}$$

and

$$\int_0^\infty \psi_j^n L(\psi_j^n)\,dx \to \int_0^\infty \phi_j L(\phi_j)\,dx. \tag{82}$$

It is clear that the R_j^n, for $j = 0$ and 1, may be replaced by $R_n = \max(R_0^n, R_1^n)$. From (81), it is seen that

$$\int_0^\infty \psi_0^n\psi_1^n dx \to \int_0^\infty \phi_0\phi_1 dx = 0, \qquad (n \to \infty). \tag{83}$$

Since

$$-\int_0^\infty \psi_0{}^n L(\psi_1{}^n)\,dx = \psi_0{}^n(0)\psi_1{}^{n\prime}(0) + \int_0^\infty (\psi_0{}^{n\prime}\psi_1{}^{n\prime} + q\psi_0{}^n\psi_1{}^n)\,dx$$

and

$$-\int_0^\infty \phi_0 L(\phi_1)\,dx = \phi_0(0)\phi_1{}'(0) + \int_0^\infty (\phi_0{}'\phi_1{}' + q\phi_0\phi_1)\,dx,$$

it follows from (80), (81) and the Schwarz inequality that

$$(84) \qquad \int_0^\infty \psi_0{}^n L(\psi_1{}^n)\,dx \to \int_0^\infty \phi_0 L(\phi_1)\,dx = 0, \qquad (n \to \infty).$$

For each n, there exist two numbers $a_0{}^n$ and $a_1{}^n$ such that

$$(85) \qquad \int_0^\infty y_n{}^2 dx = 1$$

and

$$(86) \qquad \int_0^\infty y_n(x)\phi_0(x, R_n)\,dx = 0,$$

where y_n is defined by

$$y_n = a_0{}^n\psi_0{}^n + a_1{}^n\psi_1{}^n.$$

From (81), (83) and (85) it follows that

$$(87) \qquad (a_0{}^n)^2 + (a_1{}^n)^2 \to 1, \qquad (n \to \infty).$$

From (82), (84) and (87) it is clear that

$$(88) \qquad \left| \int_0^\infty y_n L(y_n)\,dx \right| \leqq (1 + \epsilon_n)\lambda_1, \qquad 0 < \epsilon_n \to 0 \text{ as } n \to \infty.$$

But, in virtue of (86),

$$\left| \int_0^\infty y_n L(y_n)\,dx \right| \geqq \lambda_1(R_n),$$

and this last relation is in contradiction with (88) and the inequality $\lambda_1 < \lambda_1(\infty)$, provided n is sufficiently large. Consequently, relation (78) holds.

This completes the proof of Theorem (3).

If $\lambda_0 < \lambda_1 < \cdots < \lambda_n$ are isolated and are the only points λ of the spectrum S_α satisfying $\lambda_0 \leqq \lambda \leqq \lambda_n$, a theorem analogous to the preceding one holds for λ_{n+1}, the next point of the spectrum greater than λ_n. In case $q(x) \to \infty$, as $x \to \infty$, the spectrum S_α consists of a sequence (Weyl [9], p. 252) $\lambda_0 < \lambda_1 < \lambda_2 < \cdots$, where $\lambda_n \to \infty$ as $n \to \infty$, and an analogue of the theorem of Milne that $\lambda_k(X) \to \lambda_k$ as $X \to \infty$, for $k = 1, 2, \cdots$, is

obtained. In case λ_0 is a cluster point of S_a, it is clear from the proof of Theorem (3) that $\lambda_0(X), \lambda_1(X), \cdots$ all tend to λ_0 as $X \to \infty$.

10. In this section there will be proved two theorems relating the spectrum S_a with the function $q(x)$ appearing in the differential equation (3). First, there will be proved the

THEOREM (4). *Let $q = q(x)$, where $0 \leqq x < \infty$, be continuous and satisfy* (6) *and let λ_0 denote the least point of the spectrum of the boundary value problem determined by the differential equation* (3) *and the boundary condition* (1). *If $q_1(x)$, $q_2(x), \cdots$ is a sequence of continuous functions tending uniformly on the half-line $0 \leqq x < \infty$ to $q(x)$, then the least point λ_0^n of the spectrum of the boundary value problem determined by the differential equation*

$$L_n(y) + \lambda = 0, \qquad L_n(y) = y'' - q_n y, \tag{89}$$

and the boundary condition (1) *satisfies*

$$\lambda_0^n \to \lambda_0, \qquad (n \to \infty). \tag{90}$$

Proof of Theorem (4). It will be clear from the proof that there is no loss of generality in assuming that λ_0^n and λ_0 are eigenvalues with respective normalized eigenfunctions ϕ_0^n and ϕ_0 of their corresponding boundary value problems. Suppose $| q - q_n | < \epsilon_n$ for $0 \leqq x < \infty$ and consider the relation

$$| \int_0^\infty (\phi_0 L(\phi_0) - \phi_0 L_n(\phi_0))\, dx | = | \int_0^\infty (q - q_n)\phi_0^2 dx | < \epsilon_n, \tag{91}$$

and the corresponding relation where ϕ_0 is replaced by ϕ_0^n. If the integrals involving the operators L and L_n are replaced by the corresponding "spectral" integrals taken over the λ-axis, the above relations imply $\lambda_0^n < \lambda_0 + \epsilon_n$ and $\lambda_0 < \lambda_0^n + \epsilon_n$ respectively, that is, $| \lambda - \lambda_0^n | < \epsilon_n$, and the proof is complete.

It is clear from (91) that if the continuous functions q_n satisfy (6) and if the assumption of uniform convergence of the theorem is replaced by the condition

$$\int_0^\infty | q - q_n |^p dx \to 0, \qquad (n \to \infty),$$

for some $p > 1$, then (90) is still valid. For ϕ_0 and ϕ_0^n, $n = 1, 2, \cdots$, tend to 0 as x tends to ∞ (cf. **4**) and since these functions are of class (L^2) they are also of class (L^{2r}), $r > 1$. An application of the Hölder inequality to the second integral of (91) then yields the desired result.

The last theorem of this paper is

THEOREM (5). *If* $q = q(x)$, *where* $0 \leqq x < \infty$, *is continuous and satisfies relation* (4), *then* $\lambda = c$ *belongs to the spectrum* S_α *for every* α.

Proof of Theorem (5). Since the spectrum S_α is displaced by k if the function q is replaced by $q + k$, where k is any real constant, it is clear that it can be supposed that $c = 0$. If there exist an infinity of negative eigenvalues, they cluster at $\lambda = 0$ (Weyl [9], p. 252) and the theorem is proved. Suppose that there are only a finite number of negative eigenvalues and that $\lambda = 0$ is not a cluster point of S_α. Let $\mu > 0$ denote the least positive point of the spectrum and let $\phi_1, \phi_2, \cdots, \phi_n$ denote the eigenfunctions belonging to the finite number of eigenvalues $\lambda_j < \mu$ of S_α. Let ϵ be a positive number satisfying

(92) $$0 < \epsilon < \tfrac{1}{2}\mu$$

and let X be chosen so large that

(93) $$|q(x)| < \epsilon \text{ if } x \geqq X.$$

Choose $n + 2$ points $X = X_0 < X_1 < \cdots < X_{n+1}$ and functions $y_i(x)$, $i = 1, 2, \cdots, n + 1$, of class Ω_0 such that

$$y_i(x) \equiv 0 \text{ if } 0 \leqq x \leqq X_{i-1}, \qquad X_i \leqq x < \infty,$$

$$\int_0^\infty y_i^2 dx = 1$$

and

$$\int_0^\infty y_i'^2 dx < \epsilon.$$

It is clear that the $n + 1$ functions thus defined constitute an orthonormal set and that consequently there exist constants $c_1, \cdots, c_{n+1}$ for which the function

$$y = c_1 y_1 + \cdots + c_{n+1} y_{n+1}$$

satisfies the $n + 1$ conditions

$$\int_0^\infty y^2 dx = c_1^2 + \cdots + c_{n+1}^2 = 1$$

and

$$\int_0^\infty y\phi_i dx = 0, \qquad i = 1, 2, \cdots, n.$$

In virtue of (19), (57) and the preceding two relations, it follows that

$$\int_0^\infty (y'^2 + qy^2)\,dx \geqq \mu. \tag{94}$$

But

$$\int_0^\infty (y'^2 + qy^2)\,dx = \sum_{i=1}^{n+1} c_i^2 \int_0^\infty (y_i'^2 + qy_i^2)\,dx < 2\epsilon,$$

by (93) and the properties of the functions $y_i(x)$. This last relation is in contradiction with (94), in virtue of (92). This completes the proof of Theorem (5).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] T. Carleman, *Sur les équations intégrales singulières à noyau réel et symétrique*, Uppsala, 1923.

[2] G. Hardy, J. Littlewood and G. Pólya, *Inequalities*, Cambridge, 1934.

[3] P. Hartman, "The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[4] P. Hartman, "On differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[5] E. Hellinger, *Die Orthogonalinvarianten quadratischer Formen von unendlichen Variabelen (Inaugural-Dissertation)*, Göttingen, 1907.

[6] E. Hellinger, "Neue Begründung der Theorie quadratischer Formen von unendlichen Veränderlichen," *Journal für die reine und angewandte Mathematik*, vol. 136 (1909), pp. 210-271.

[7] L. Lichtenstein, "Zur Variationsrechnung," *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen* (1919), pp. 161-192.

[8] W. Milne, "The behavior of a boundary value problem as the interval becomes infinite," *Transactions of the American Mathematical Society*, vol. 30 (1928), pp. 797-802.

[9] H. Weyl, "Über gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

[10] H. Weyl, *Gruppentheorie und Quantenmechanik*, Leipzig, 1928.

[11] A. Wintner, *Spektraltheorie der unendlichen Matrizen*, Leipzig, 1929.

[12] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

TRANSFORMATIONS OF DOUBLE SEQUENCES.*

By Florence M. Mears.

1. **Introduction.** Let $\| a_{mnkl} \|$ be a matrix of real or complex constants defining a transformation

$$U'_{mn} = \sum_{k,l=0}^{\infty} a_{mnkl} U_{kl},$$

by means of which the double sequence $\{U_{kl}\}$ is transformed into the double sequence $\{U'_{mn}\}$. Let $u_{pq} = U_{pq} - U_{p,q-1} - U_{p-1,q} + U_{p-1,q-1}$, with $U_{p,-1}$, $U_{-1,q}$ and $U_{-1,-1}$ defined as zero; then $U_{kl} = \sum_{p,q=0}^{k,l} u_{pq}$. We define u'_{ij} in a similar manner in terms of U'_{mn} so that $U'_{mn} = \sum_{i,j=0}^{m,n} u'_{ij}$.

Let $\| b_{mnkl} \|$ be a matrix defined by the equation $b_{mnkl} = a_{mnkl} - a_{m,n-1,kl} - a_{m-1,nkl} + a_{m-1,n-1,kl}$, with $a_{m,-1,kl}$, $a_{-1,nkl}$ and $a_{-1,-1,kl}$ defined as zero; then $a_{mnkl} = \sum_{i,j=0}^{m,n} b_{ijkl}$. This matrix defines a transformation

$$u'_{mn} = \sum_{k,l=0}^{\infty} b_{mnkl} U_{kl},$$

by means of which the double sequence $\{U_{kl}\}$ is transformed into the double series $\sum_{m,n=0}^{\infty} u'_{mn}$.

In two papers published in 1936, Hamilton ([2] and [3]) made use of the following definitions and abbreviations. The sequence $\{U_{kl}\}$ is said to be existent, *e*, if U_{kl} exists, $k, l = 0, 1, \cdots$; ultimately bounded, *ub*, if there exists a number Q_1 such that U_{kl} is bounded for all $k, l > Q_1$; bounded, *b*, if *ub* with $Q_1 = 0$; convergent, *c*, if there exists a number U such that $\lim_{k,l\to\infty} U_{kl} = U$; bounded convergent, *bc*, if *b* and *c*; ultimately regularly convergent, *urc*, if *c*, and if there exist numbers Q_2, $U_{k\infty}$ and $U_{\infty l}$ such that $\lim_{l\to\infty} U_{kl} = U_{k\infty}$ for all $k > Q_2$ and $\lim_{k\to\infty} U_{kl} = U_{\infty l}$ for all $l > Q_2$; regularly convergent, *rc*, if *urc* with $Q_2 = 0$; bounded ultimately regularly convergent, *burc*, if *b* and *urc*; *cn*, if *c* and $U = 0$; *bcn*, if *bc* and $U = 0$; *urcn*, if *urc*

* Received October 9, 1947.

and $U=0$; *rcn*, if *rc* and $U=0$; *burcn*, if *burc* and $U=0$; *urcrn*, if *urc* and if there exists a number Q_3 such that $U_{k\infty}=0$, $k>Q_3$, $U_{\infty l}=0$, $l>Q_3$; *rcurn*, if *rc* and $U_{k\infty}=0$, $k>Q_3$, $U_{\infty l}=0$, $l>Q_3$; *burcrn*, if *burc* and $U_{k\infty}=0$, $k>Q_3$ and $U_{\infty l}=0$, $l>Q_3$; row null, *rcrn*, if *rcurn* and $Q_3=0$.

References and a diagram showing the relations among the various types may be obtained from [3].

Included among the results obtained by Hamilton [3] are theorems stating the conditions which must be satisfied by the matrix $\| a_{mnkl} \|$ in order that a sequence, $\{U_{kl}\}$, which is any specific one of the last sixteen types listed here, have an existent transform, $\{U'_{mn}\}$, which is also a specific one of these types; conditions sufficient for the same purpose; necessary and sufficient conditions for convergence preservation, with $U'=0$ when $U=0$, and necessary and sufficient conditions for regularity, that is, for convergence preservation with $U'=U$.

In the present paper we shall consider also the following types of sequences: absolutely convergent, *a*, if there exists a number U^* such that $\sum_{k,l=0}^{\infty} | u_{kl} | = U^*$; absolutely convergent null, *an*, if *a* and *cn*; *aurn*, if *a* and *urcrn*; *arn*, if *a* and *rcrn*. It is obvious that *a* implies *rc*, *an* implies *rcn*, and *aurn* implies *rcurn*. We shall denote $\lim_{m,n\to\infty} U'_{mn}$, $\lim_{n\to\infty} U'_{mn}$ and $\lim_{m\to\infty} U'_{mn}$ by U', $U'_{m\infty}$ and $U'_{\infty n}$, respectively, if these limits exist.

We shall obtain theorems corresponding to those obtained by Hamilton [3], and described above, when $\{U_{kl}\}$ is any specific one of the last four types, and $\{U'_{mn}\}$ any one of the last twenty types, listed above.

For the purpose of comparison, we shall present briefly, in Section **2**, similar theorems for simple sequences. The remainder of the paper is concerned with double sequences; in Section **3** are listed the conditions to be used; various implications of these conditions are given in Section **4**, necessity theorems in Section **5**, sufficiency theorems in Section **6**, and a few concluding remarks in Section **7**.

2. Simple sequences. Let $\| a_{nk} \|$ be a matrix of real or complex constants defining a transformation $U'_n=\sum_{k=0}^{\infty} a_{nk}U_k$. Let $u_k=U_k-U_{k-1}$ and $u'_n=U'_n-U'_{n-1}$ (with $U_{-1}=U'_{-1}=0$). Of the types considered for double sequences in Section 1, only *e*, *b*, *c*, *cn*, *a* and *an* are applicable to simple sequences; their definitions for simple sequences are obvious.

We are concerned with theorems which involve transformations of sequences $\{U_k\}$, of type *a* or *an*, into existent sequences $\{U'_n\}$, of type *b*, *c*,

cn, a or an. The theorems in which $\{U'_n\}$ is b or c have been proved by Hahn [1]; the theorem in which $\{U_k\}$ and $\{U'_n\}$ are both a has been proved by the writer [4]. The additional conditions necessary and sufficient for the remaining theorems are easily obtained.

We shall use the following conditions:

$$(1) \qquad \Big|\sum_{k=0}^{p} a_{nk}\Big| < B_1(n); \qquad p, n = 0, 1, \cdots;$$

$$(2) \qquad \sum_{k=0}^{\infty} a_{nk} = A_n; \qquad n = 0, 1, \cdots;$$

$$(3) \qquad (1) \text{ and } B_1(n) < B_1; \qquad n = 0, 1, \cdots;$$

$$(4) \qquad \lim_{n\to\infty} a_{nk} = \alpha_k; \qquad k = 0, 1, \cdots;$$

$$(5) \qquad (2) \text{ and } \lim_{n\to\infty} A_n = A;$$

$$(6) \qquad (4) \text{ and } \alpha_k = 0; \qquad k = 0, 1, \cdots;$$

$$(7) \qquad (5) \text{ and } A = 0;$$

$$(8) \qquad \sum_{n=0}^{\infty}\Big|\sum_{k=0}^{p}(a_{nk} - a_{n-1,k})\Big| < B_2; \qquad p = 0, 1, \cdots.$$

It is easily proved that (8) implies (3) and (4), and that (2) + (8) implies (5).

In the list below, the abbreviation "$x \to y(a)$" is to be read "in order that every sequence $\{U_k\}$ of type x have a transform $\{U'_n\}$ of type y, it is necessary and sufficient that $\| a_{nk} \|$ satisfy (a)."

$an \to e$	(1).	
$a \to e$	(2).	
$an \to b$	(3).	([1], p. 26, Th. XIa)
$a \to b$	(2) and (3).	([1], p. 24, Th. Xa)
$an \to c$	(3) and (4).	([1], p. 27, Th. XIb)
$a \to c$	(3), (4), and (5).	([1], p. 25, Th. Xb)
$an \to cn$	(3) and (6).	
$a \to cn$	(3), (6) and (7).	

$an \to a$ (8).

$a \to a$ (2) and (8). ([4], p. 595)

$an \to an$ (6) and (8).

$a \to an$ (6), (7) and (8).

If $\{U_k\}$ is a and $\{U'_n\}$ a or c, and if $\lim_{k\to\infty} U_k = U$, we have $\lim_{n\to\infty} U'_n = \sum_{k=0}^{\infty} \alpha_k(U_k - U) + AU$. In order that $\lim_{n\to\infty} U'_n = AU$, it is sufficient that $\alpha_k = 0$; therefore convergence is preserved, with preservation of the limit for null sequences, if, in the theorems stating the necessary and sufficient conditions for these transformations, condition (4) (which is included by implication in the case of $a \to a$) is replaced by (6). If, in addition, $A = 1$, the transformation is regular. In both cases the conditions are necessary.

3. Conditions on the matrix. We shall make use of the following twenty-six conditions. Twenty of them are identical (except for notation) with the two-dimensional form of specific conditions used by Hamilton ([3], pp. 35-37). When this relationship exists, the statement of the condition is preceded by two designations; the first in the pair is the listing which will be used throughout this paper, the second that used by Hamilton.

(9) $\left|\sum_{k,l=0}^{p,q} a_{mnkl}\right| < B_1(m,n)$; $p, q, m, n = 0, 1, \cdots$;

(10) $\sum_{l=0}^{\infty} a_{mnkl} = \alpha_{mnk\infty}$; $k, m, n = 0, 1, \cdots$;

$\sum_{k=0}^{\infty} a_{mnkl} = \alpha_{mn\infty l}$; $l, m, n = 0, 1, \cdots$;

(11) $\sum_{k,l=0}^{\infty} a_{mnkl} = A_{mn}$; $m, n = 0, 1, \cdots$;

(12) (9) and $B_1(m,n) < B_1$; $m, n \geq M_1$;

(13) (12) with $M_1 = 0$;

(14) (d_1) $\lim_{m,n\to\infty} a_{mnkl} = a_{kl}$; $k, l = 0, 1, \cdots$;

(15) (d_2) (10) and $\lim_{m,n\to\infty} \alpha_{mnk\infty} = \alpha_{k\infty}$; $k = 0, 1, \cdots$;

$\lim_{m,n\to\infty} \alpha_{mn\infty l} = \alpha_{\infty l}$; $l = 0, 1, \cdots$;

(16) (d_3) (11) and $\lim\limits_{m,n\to\infty} A_{mn} = A$;

(17) ($\bar{d}_1$) (14) with $a_{kl} = 0$; $k, l = 0, 1, \cdots$;

(18) ($\bar{d}_2$) (15) with $\alpha_{k\infty} = \alpha_{\infty l} = 0$; $k, l = 0, 1, \cdots$;

(19) ($\bar{d}_3$) (16) with $A = 0$;

(20) (e_1) $\lim\limits_{n\to\infty} a_{mnkl} = a_{m\infty kl}$; $m \geqq M_2$; $k, l = 0, 1, \cdots$;

$\lim\limits_{m\to\infty} a_{mnkl} = a_{\infty nkl}$; $n \geqq M_2$; $k, l = 0, 1, \cdots$;

(21) (e^*_2) (10) and $\lim\limits_{n\to\infty} \alpha_{mnk\infty} = \alpha_{m\infty k\infty}$; $m \geqq M_3(k)$; $k = 0, 1, \cdots$;

$\lim\limits_{m\to\infty} \alpha_{mnk\infty} = \alpha_{\infty nk\infty}$; $n \geqq M_3(k)$; $k = 0, 1, \cdots$;

$\lim\limits_{n\to\infty} \alpha_{mn\infty l} = \alpha_{m\infty\infty l}$; $m \geqq M_3(l)$; $l = 0, 1, \cdots$;

$\lim\limits_{m\to\infty} \alpha_{mn\infty l} = \alpha_{\infty n\infty l}$; $n \geqq M_3(l)$; $l = 0, 1, \cdots$;

(22) (e_2) (21) with $M_3(k), M_3(l) \leqq M_3$;

(23) (e_3) (11) and $\lim\limits_{n\to\infty} A_{mn} = A_{m\infty}$; $m \geqq M_4$;

$\lim\limits_{m\to\infty} A_{mn} = A_{\infty n}$; $n \geqq M_4$;

(24) ($\bar{e}_1$) (20) with $a_{m\infty kl} = 0$; $m \geqq M_5$; $k, l = 0, 1, \cdots$;

$a_{\infty nkl} = 0$; $n \geqq M_5$; $k, l = 0, 1, \cdots$;

(25) ($\bar{e}^*_2$) (21) with $\alpha_{m\infty k\infty} = 0$; $m \geqq M_6(k)$; $k = 0, 1, \cdots$;

$\alpha_{\infty nk\infty} = 0$; $n \geqq M_6(k)$; $k = 0, 1, \cdots$;

$\alpha_{m\infty\infty l} = 0$; $m \geqq M_6(l)$; $l = 0, 1, \cdots$;

$\alpha_{\infty n\infty l} = 0$; $n \geqq M_6(l)$; $l = 0, 1, \cdots$;

(26) ($\bar{e}_2$) (25) with $M_6(k), M_6(l) \leqq M_6$;

(27) ($\bar{e}_3$) (23) with $A_{m\infty} = 0$; $m \geqq M_7$;

$A_{\infty n} = 0$; $n \geqq M_7$;

(28) (f_1) (20) with $M_2 = 0$;

(29) (f_2) (22) with $M_3 = 0$;

(30) (f_3) (23) with $M_4 = 0$;

(31) ($\bar{f}_1$) (28) with $a_{m\infty kl} = a_{\infty nkl} = 0$;

(32) ($\bar{f}_2$) (29) with $\alpha_{m\infty k\infty} = \alpha_{\infty nk\infty} = \alpha_{m\infty\infty l} = \alpha_{\infty n\infty l} = 0$;

(33) ($\bar{f}_3$) (30) with $A_{m\infty} = A_{\infty n} = 0$;

(34) $$\sum_{m,n=0}^{\infty} \left| \sum_{k,l=0}^{p,q} b_{mnkl} \right| < B_2; \qquad p, q = 0, 1, \cdots.$$

4. Implications of conditions. Throughout the remainder of the paper, proofs are given or indicated unless they seem obvious. The symbol $\rightarrow$ denotes implication.

(35) (9) is equivalent to $\left| \sum_{k=p,\, l=q}^{p+r,\, q+s} a_{mnkl} \right| < B_1(m, n)$, $p, q, r, s, m, n = 0, 1, \cdots$.

Proof. The existence of an upper bound for $\left| \sum_{k,l=0}^{p,q} a_{mnkl} \right|$ implies the existence of an upper bound for $\left| \sum_{k=p,\, l=q}^{p+r,\, q+s} a_{mnkl} \right|$ since $\left| \sum_{k=p,\, l=q}^{p+r,\, q+s} a_{mnkl} \right| = \left| \sum_{k,l=0}^{p+r,\, q+s} a_{mnkl} - \sum_{k,l=0}^{p+r,\, q-1} a_{mnkl} - \sum_{k,l=0}^{p-1,\, q+s} a_{mnkl} + \sum_{k,l=0}^{p-1,\, q-1} a_{mnkl} \right|$. The converse is obvious.

(36) (10) + (11) $\rightarrow$ (9).

(37) $$(12) + (14) \rightarrow \left| \sum_{k,l=0}^{p,q} a_{kl} \right| \leq B_1; \qquad p, q = 0, 1, \cdots.$$

Proof. For arbitrarily chosen p and q and for a given $\epsilon > 0$,

$$\left| \sum_{k,l=0}^{p,q} a_{kl} \right| < \left| \sum_{k,l=0}^{p,q} a_{mnkl} \right| + \epsilon < B_1 + \epsilon,$$

for m and n sufficiently large.

(38) $$(12) + (15) \rightarrow \left| \sum_{k=0}^{p} \alpha_{k\infty} \right| \leq B_1; \qquad p = 0, 1, \cdots;$$

$$\left| \sum_{l=0}^{q} \alpha_{\infty l} \right| \leq B_1; \qquad q = 0, 1, \cdots.$$

Proof. For arbitrarily chosen p and for a given $\epsilon > 0$,

$$\left| \sum_{k=0}^{p} \alpha_{k\infty} \right| < \left| \sum_{k=0}^{p} \alpha_{mnk\infty} \right| + \epsilon < B_1 + \epsilon,$$

for m and n sufficiently large.

(39) $\quad (12) + (20) \rightarrow \left| \sum_{k,l=0}^{p,q} a_{m\infty kl} \right| \leq B_1;$
$$p, q = 0, 1, \cdots; m \geq M_1 + M_2;$$
$$\left| \sum_{k,l=0}^{p,q} a_{\infty nkl} \right| \leq B_1;$$
$$p, q = 0, 1, \cdots; n \geq M_1 + M_2.$$

(40) $\quad (12) + (22) \rightarrow \left| \sum_{k=0}^{p} \alpha_{m\infty k\infty} \right| \leq B_1;$
$$p = 0, 1, \cdots; m \geq M_1 + M_3;$$
$$\left| \sum_{k=0}^{p} \alpha_{\infty nk\infty} \right| \leq B_1;$$
$$p = 0, 1, \cdots; n \geq M_1 + M_3;$$
$$\left| \sum_{l=0}^{q} \alpha_{m\infty\infty l} \right| \leq B_1;$$
$$q = 0, 1, \cdots; m \geq M_1 + M_3;$$
$$\left| \sum_{l=0}^{q} \alpha_{\infty n\infty l} \right| \leq B_1;$$
$$q = 0, 1, \cdots; n \geq M_1 + M_3.$$

(41) $\quad (14) + (24) \rightarrow (17)$. $\qquad$ ([3], p. 40, relation .73)

(42) $\quad$ (34) is equivalent to $\sum_{m,n=0}^{\infty} \left| \sum_{k=p,\, l=q}^{p+r,\, q+s} b_{mnkl} \right| < B_2;$
$$p, q, r, s = 0, 1, \cdots.$$

Proof. See the proof of (35).

(43) $\quad (13) + (28) \rightarrow$ the inequalities in (39) with $M_1 + M_2 = 0$.

(44) $\quad (13) + (29) \rightarrow$ the inequalities in (40) with $M_1 + M_3 = 0$.

(45) $\quad (34) \rightarrow (13)$.

Proof. For arbitrarily chosen m_0 and n_0, we have $\left| \sum_{k,l=0}^{p,q} a_{m_0 n_0 kl} \right|$ $= \left| \sum_{k,l=0}^{p,q} \sum_{m,n=0}^{m_0,n_0} b_{mnkl} \right| = \left| \sum_{m,n=0}^{m_0,n_0} \sum_{k,l=0}^{p,q} b_{mnkl} \right| \leq \sum_{m,n=0}^{\infty} \left| \sum_{k,l=0}^{p,q} b_{mnkl} \right| < B_2$, by (34); therefore (13) is satisfied.

(46) $\quad (34) \rightarrow (14)$.

Proof. Using (42), with $r = s = 0$, we have $\sum_{m,n=0}^{\infty} | b_{mnpq} | < B_2$, p, q

$=0,1,\cdots$, therefore $\sum_{m,n=0}^{\infty} b_{mnpq}$ converges. Since $a_{mnkl}=\sum_{i,j=0}^{m,n} b_{ijkl}$, it follows that $\lim_{m,n\to\infty} a_{mnkl}$ exists, $k,l=0,1,\cdots$.

(47) (34) $\to$ (20).

Proof. Using the proof of (46), we see that $\sum_{m,n=0}^{\infty} b_{mnpq}$ converges absolutely, $p,q=0,1,\cdots$; therefore for any arbitrarily chosen m, $\lim_{n\to\infty}\sum_{i,j=0}^{m,n} b_{ijkl}$ exists.

(48) (34) $\to$ (28).

Proof. See proof of (46).

(49) (10) + (34) $\to$ (15).

Proof. Since (10) is satisfied, $\sum_{l=0}^{\infty} b_{mnkl}$ converges; $m,n,k=0,1,\cdots$; therefore for arbitrarily chosen m_0, n_0 and k_0, we can find q_0 so that $\left|\sum_{l=q_0+1}^{\infty} b_{mnk_0l}\right| < B_2(m_0+1)^{-1}(n_0+1)^{-1}$ for $m=0,1,\cdots,m_0$ and $n=0,1,\cdots,n_0$. Using this and (34) we have $\sum_{m,n=0}^{m_0,n_0}\left|\sum_{l=0}^{\infty} b_{mnk_0l}\right| = \sum_{m,n=0}^{m_0,n_0}\left|\sum_{l=0}^{q_0} b_{mnk_0l} + \sum_{l=q_0+1}^{\infty} b_{mnk_0l}\right| < 2B_2$. Therefore $\sum_{m,n=0}^{\infty}\left|\sum_{l=0}^{\infty} b_{mnkl}\right|$ converges, $k=0,1,\cdots$; since $\alpha_{mnk\infty}=\sum_{l=0}^{\infty}\sum_{i,j=0}^{m,n} b_{ijkl}=\sum_{i,j=0}^{m,n}\sum_{l=0}^{\infty} b_{ijkl}$, $\lim_{m,n\to\infty}\alpha_{mnk\infty}$ exists.

(50) (11) + (34) $\to$ (16).

Proof. Since (11) is satisfied, $\sum_{k,l=0}^{\infty} b_{mnkl}$ converges, $m,n=0,1,\cdots$. By a proof similar to that of (49), we find that $\sum_{m,n=0}^{\infty}\left|\sum_{k,l=0}^{\infty} b_{mnkl}\right|$ converges, and hence that $\lim_{m,n\to\infty} A_{mn}$ exists.

(51) (10) + (34) $\to$ (29).

Proof. Using the proof of (49), with m_0 and k_0 chosen arbitrarily, we have $\sum_{n=0}^{\infty}\left|\sum_{l=0}^{\infty}\sum_{m=0}^{m_0} b_{mnk_0l}\right|$ convergent. Since $\sum_{l=0}^{\infty}\sum_{m=0}^{m_0} b_{mnk_0l}=\alpha_{m_0nk_0\infty}-\alpha_{m_0,n-1,k_0\infty}$, $\lim_{n\to\infty}\alpha_{m_0nk_0\infty}$ exists, $m_0,k_0=0,1,\cdots$.

(52) $\quad (11) + (34) \rightarrow (30)$.

Proof. For arbitrarily chosen m_0, the convergence of $\sum_{n=0}^{\infty} \left| \sum_{k,l=0}^{\infty} \sum_{m=0}^{m_0} b_{mnkl} \right|$ follows from the proof of (50); therefore (30) is satisfied since $A_{m_0 n} - A_{m_0, n-1} = \sum_{k,l=0}^{\infty} \sum_{m=0}^{m_0} b_{mnkl}$.

5. Necessity theorems. Let x and y represent any of the types defined in Section **1**, and let (a) represent any condition listed in Section **3**. We shall abbreviate "in order that every sequence $\{U_{kl}\}$ of type x have a transform $\{U'_{mn}\}$ of type y, it is necessary that $\| a_{mnkl} \|$ satisfy (a)" as follows: $N \cdot x \rightarrow y$ is (a).

Let x, y, X and Y be types of sequences such that $X \supset x$ and $Y \supset y$; it is obvious that if condition (a) is $N \cdot x \rightarrow Y$ it is also $N \cdot x \rightarrow y$, $N \cdot X \rightarrow y$ and $N \cdot X \rightarrow Y$. Therefore from the necessity theorems given in this section we can obtain conditions which are necessary in order that a sequence of any one of the types *arn*, *aurn*, *an* or *a* have a transform which is a specific one of the types listed in Section **1**. For example, if $y = urcn$, Y may be *e*, *ub*, *c*, *cn*, *urc* or *urcn*; therefore we find from Theorems 1, 4, 6, 9, 12 that $N \cdot arn \rightarrow urcn$ are (9), (12), (14), (17), (20). Since $(12) \rightarrow (9)$ and since $(17) \rightarrow (14)$, we may write: $N \cdot arn \rightarrow urcn$ are (12), (17) and (20). As a second example, let $X = a$ and $y = b$; then x may be *arn*, *aurn*, *an* or *a*, and Y may be *e*, *ub* or *b*. Therefore we find from Theorems 1, 2, 3, 4, 5 that $N \cdot a \rightarrow b$ are (9), (10), (11), (12), (13), and using (36) and the fact that $(13) \rightarrow (12)$ we may write $N \cdot a \rightarrow b$ are (10), (11) and (13).

THEOREM 1. *$N \cdot arn \rightarrow e$ is (9).*

Proof. Let $A_i(m, n) = \sum_{k=p_i,\, l=q_i}^{p_i+r_i,\, q_i+s_i} a_{mnkl}$. By denial of (9), using (35), we see that there exist constants m_0 and n_0 such that $\{| A_i(m_0, n_0) |\}$ diverges. It is obvious that either (a) $p_i + r_i$ or $q_i + s_i$ (but not both), may be taken to be bounded or (b) neither may be so taken. If $p_i + r_i$ is bounded, we can choose $p_i = p_0$, $r_i = r_0$ and $q_{i+1} > q_i + s_i$ so that $| A_i(m_0, n_0) | > 2^i$, $i = 0, 1, \cdots$; if the divergence requires that $p_i + r_i \rightarrow \infty$ and $q_i + s_i \rightarrow \infty$, we can choose $p_{i+1} > p_i + r_i$ and $q_{i+1} > q_i + s_i$ so that $| A_i(m_0, n_0) | > 2^i$, $i = 0, 1, \cdots$. Let $U_{kl} = 2^{-i} \operatorname{sgn} A_i(m_0, n_0)$, $k = p_i, \cdots, p_i + r_i$, $l = q_i, \cdots, q_i + s_i$, $i = 0, 1, \cdots$; $U_{kl} = 0$, otherwise. Then $\sum_{k,l=0}^{\infty} | u_{kl} | \leq 4 \sum_{i=0}^{\infty} 2^{-i}$, and $U_{k\infty} = U_{\infty l} = 0$, $k, l = 0, 1, \cdots$; therefore $\{U_{kl}\}$ is *arn*. Let

$$S_i = \left| \sum_{k,l=0}^{p+r,q_i+s_i} a_{m_0n_0kl}U_{kl} - \sum_{k,l=0}^{p-1,q_i-1} a_{m_0n_0kl}U_{kl} \right|.$$

If $p_i + r_i$ is bounded, $S_i = |A_i 2^{-i}| > 1$, for $p > p_0 + r_0$, $i = 0, 1, \cdots$; if $p_i + r_i \to \infty$ and $q_i + s_i \to \infty$, $S_i = |A_i 2^{-i}| > 1$, $p = p_i$, $r = r_i$, $i = 0, 1, \cdots$. Therefore $U'_{m_0n_0}$ does not exist.

THEOREM 2. *$N \cdot aurn \to e$ is* (10).

Proof. By denial of (10) for some m_0, n_0 and k_0 $\sum_{l=0}^{\infty} a_{m_0n_0k_0l}$ does not converge. Let $U_{kl} = 1$ for $k = k_0$ and $l = 0, 1, \cdots$; $U_{kl} = 0$, otherwise. Then $\{U_{kl}\}$ is *aurn* but $U'_{m_0n_0}$ does not exist.

THEOREM 3. *$N \cdot a \to e$ is* (11).

Proof. By denial of (11) for some m_0 and n_0 $\sum_{k,l=0}^{\infty} a_{m_0n_0kl}$ does not exist. Let $U_{kl} = 1$, $k, l = 0, 1, \cdots$. Then $\{U_{kl}\}$ is *a*, but $U'_{m_0n_0}$ does not exist.

THEOREM 4. *$N \cdot arn \to ub$ is* (12).

Proof. Let $A_i(m, n)$ be defined as in Theorem 1. By denial of (12), there exists a sequence $\{|A_i(m_i, n_i)|\}$ which, with m_i and n_i, becomes infinite with i.

(a) If $\{p_i + r_i\}$, $\{q_i + s_i\}$ may both be taken to be bounded, we can choose $p_i = p_0$, $r_i = r_0$, $q_i = q_0$, $s_i = s_0$ and $U_{kl} = 1$, $k = p_0, \cdots, p_0 + r_0$, $l = q_0, \cdots, q_0 + s_0$; $U_{kl} = 0$, otherwise. Then

$$\overline{\lim_{m,n\to\infty}} |U'_{mn}| = \lim_{i\to\infty} |A_i(m_i, n_i)| = \infty.$$

(b) From the preceding paragraph it follows that there exist constants $C(i)$, $i = 0, 1, \cdots$, such that $|A_i(m, n)| < C(i)$, $m, n = 0, 1, \cdots$; since (9) is $N \cdot arn \to e$, we may assume (9) satisfied and we have $|A_i(m, n)| < B_1(m, n)$, $i = 0, 1, \cdots$. If $\{p_i + r_i\}$ is bounded, but $\{q_i + s_i\}$ necessarily unbounded, we can choose $p_i = p_0$, $r_i = r_0$ and m_0, n_0, q_0 and s_0 so that $|A_0(m_0, n_0)| > 1$; we can choose $q_i > q_{i-1} + s_{i-1}$, $m_i > m_{i-1}$, $n_i > n_{i-1}$, $i = 1, 2, \cdots$, so that

$$|A_i(m_i, n_i)| > 2^i B_1(m_{i-1}, n_{i-1}) [i + 1 + \sum_{j=0}^{i-1} C(j)/2^j B_1(m_{j-1}, n_{j-1})],$$

with $B_1(m_{-1}, n_{-1})$ defined as one. Let $U_{kl} = 1$, $k = p_0, \cdots, p_0 + r_0$, $l = q_0, \cdots, q_0 + s_0$; $U_{kl} = [2^i B_1(m_{i-1}, n_{i-1})]^{-1}$, $k = p_0, \cdots, p_0 + r_0$, $l = q_i, \cdots, q_i + s_i$; $U_{kl} = 0$, otherwise. Then

$$
\begin{aligned}
| U'_{m_i n_i} | \geq & | A_i(m_i, n_i) | [2^i B_1(m_{i-1}, n_{i-1})]^{-1} - | A_0(m_i, n_i) | \\
& - \sum_{j=1}^{i-1} | A_j(m_i, n_i) | [2^j B_1(m_{j-1}, n_{j-1})]^{-1} \\
& - \sum_{j=i+1}^{\infty} | A_j(m_i, n_i) | [2^j B_1(m_{j-1}, n_{j-1})]^{-1} \\
\geq & \; i + 1 - B_1(m_i, n_i) \sum_{j=i+1}^{\infty} [2^j B_1(m_{j-1}, n_{j-1})]^{-1}, \\
& \qquad\qquad\qquad\qquad\qquad i = 2, 3, \cdots,
\end{aligned}
$$

and since $B_1(m_i, n_i) \leq B_1(m_j, n_j)$, $j \geq i$, we have

$$| U'_{m_i n_i} | \geq i + 1 - \sum_{j=i+1}^{\infty} 2^{-j} \geq i.$$

(c) If necessarily both $p_i + r_i$ and $q_i + s_i \to \infty$ as $i \to \infty$, the proof is the same as that of (b), except that we choose $p_i > p_{i-1} + r_{i-1}$.

THEOREM 5. *$N \cdot arn \to b$ is* (13).

Proof. By Theorem 4, since $ub \supset b$, we may assume (12) satisfied; by denial of (13) there exists a sequence $\{| A_i(m_0, n_i) |\}$ which with n_i becomes infinite with i. In the proof of Theorem 4, let $m_i = m_0$ and replace $m, n \to \infty$ by $n \to \infty$.

THEOREM 6. *$N \cdot arn \to c$ is* (14).

Proof. Assume that $\lim_{m,n\to\infty} a_{mnk_0l_0}$ does not exist. The sequence $U_{kl} = 1$, for $k = k_0$, $l = l_0$, $U_{kl} = 0$, otherwise, used by Hamilton ([3], p. 42, Theorem 7) to prove that (14) is $N \cdot rcrn \to c$, is not only *rcrn*, but also *arn*.

THEOREM 7. *$N \cdot aurn \to c$ is* (15).

Proof. Assume that $\lim_{m,n\to\infty} \alpha_{mnk_0\infty}$ does not exist. The sequence $U_{kl} = 1$, for $k = k_0$, $l = 0, 1, \cdots$, $U_{kl} = 0$, otherwise, used by Hamilton ([3], p. 42, Theorem 8) to prove that (15) is $N \cdot rcurn \to c$, is not only *rcurn*, but also *aurn*.

THEOREM 8. *$N \cdot a \to c$ is* (16).

Proof. Assume that $\lim_{m,n\to\infty} A_{mn}$ does not exist. The sequence $U_{kl} = 1$, $k, l = 0, 1, \cdots$, used by Hamilton ([3], p. 42, Theorem 9) to prove that (16) is $N \cdot rc \to c$, is not only *rc* but also *a*.

THEOREM 9. *$N \cdot arn \to cn$ is* (17).

Proof. By Theorem 6, since $c \supset cn$, we may assume (14) satisfied; assume $a_{k_0 l_0} \neq 0$. See the proof of Theorem 6 ([3], p. 42, Theorem 12).

THEOREM 10. *$N \cdot aurn \to cn$ is* (18).

Proof. By Theorem 7, we may assume (15) satisfied; assume $\alpha_{k_0 \infty} \neq 0$. See the proof of Theorem 7 ([3], p. 43, Theorem 13).

THEOREM 11. *$N \cdot a \to cn$ is* (19).

Proof. By Theorem 8, we may assume (16) satisfied; assume $A \neq 0$. See the proof of Theorem 8 ([3], p. 43, Theorem 14).

THEOREM 12. *$N \cdot arn \to urc$ is* (20).

Proof. By Theorems 4 and 6, since $ub \supset c \supset urc$, we may assume (12) and (14) satisfied; by denial of (20) we may assume the existence of an increasing sequence $\{m_i\}$ and sequences $\{k_i\}$ and $\{l_i\}$ such that $\lim_{n\to\infty} a_{m_i n k_i l_i}$ does not exist, $i = 0, 1, \cdots$.

(a) Assume $k_i = k_0$, $l_i = l_0$; let $U_{kl} = 1$ for $k = k_0$, $l = l_0$; $U_{kl} = 0$, otherwise.

(b) From the preceding paragraph it follows that there exist constants $M_2(k,l)$ such that $\lim_{n\to\infty} a_{mnkl}$ exists, $m > M_2(k,l)$, $k, l = 0, 1, \cdots$. We can choose δ_0, m_0, N_0, k_0 and l_0 satisfying $0 < \delta_0 \leq (B_1 + 1)^{-1}$ and $m_0 > M_1$, $N_0 > M_1$, so that

$$\operatorname{osc} a_{m_0 n k_0 l_0} > \delta_0(B_1 + 1), \qquad n > N_0.$$

We can choose, for $i = 1, 2, \cdots$, δ_i, m_i, N_i, k_i and l_i, satisfying $0 < \delta_i \leq \delta_{i-1}$, $m_i > m_{i-1} + M_2(k_{i-1}, l_{i-1})$, $N_i > M_1$, $k_i \geq k_{i-1}$, $l_i \geq l_{i-1}$ (with at most one equality sign holding in the last two conditions for a given i), so that

$$\lim_{n\to\infty} a_{m_i n k_j l_j} = a_{m_i \infty k_j l_j} \qquad j = 0, 1, \cdots, i-1, \tag{53}$$

$$\operatorname{osc} a_{m_i n k_i l_i} > \delta_i(B_1 + 1), \qquad n > N_i. \tag{54}$$

Let $U_{kl} = 1$ for $k = k_0$, $l = l_0$; $U_{k_i l_i} = 4^{-i} \prod_{j=0}^{i-1} \delta_j$ for $i = 1, 2, \cdots$; $U_{kl} = 0$, otherwise. By (53),

$$\lim_{n\to\infty} \sum_{j=0}^{i-1} a_{m_i n k_j l_j} U_{k_j l_j}$$

exists. By (54), for $n > N_i$,

$$\operatorname{osc}\, a_{m_i n k_i l_i} U_{k_i l_i} > 4^{-i}(B_1 + 1) \prod_{j=0}^{i} \delta_j ;$$

by (12), for $n > N_i$,

$$\left| \sum_{j=i+1}^{\infty} a_{m_i n k_j l_j} U_{k_j l_j} \right| < (B_1 + 1) \sum_{j=i+1}^{\infty} | U_{k_j l_j} |$$

$$< 3^{-1} 4^{-i} (B_1 + 1) \prod_{j=0}^{i} \delta_j ;$$

therefore

$$\operatorname{osc} \sum_{j=i}^{\infty} a_{m_i n k_j l_j} U_{k_j l_j} > 2 \cdot 3^{-1} \cdot 4^{-i} (B_1 + 1) \prod_{j=0}^{i} \delta_j .$$

It follows that $U'_{m_i \infty}$ does not exist.

THEOREM 13. *$N \cdot aurn \to urc$ is* (21).

Proof. By denial of (21), we may assume the existence of an increasing sequence $\{m_i\}$ such that $\lim_{n \to \infty} \alpha_{m_i n k_0 \infty}$ does not exist, $i = 0, 1, \cdots$. See the proof of Theorem 7.

THEOREM 14. *$N \cdot an \to urc$ is* (22).

Proof. By Theorems 4 and 13, since $ub \supset urc$ and $an \supset aurn$, we may assume (12) and (21) satisfied. By denial of (22), we can choose an increasing sequence $\{k_i\}$, $i = 0, 1, \cdots$, so that $M_3(k_i) \to \infty$ as $i \to \infty$. We can choose δ_0, m_0 and k_0 satisfying $0 < \delta_0 \leq (B_1 + 1)^{-1}$, and $m_0 > M_1$, so that

$$\operatorname{osc}\, \alpha_{m_0 n k_0 \infty} > \delta_0 (B_1 + 1),$$

but, by (21), there exists a constant $M_3(k_0) > m_0$ such that

$$\alpha_{m \infty k_0 \infty} \text{ exists,} \qquad m \geq M_3(k_0).$$

We can choose, for $i = 1, 2, \cdots, \delta_i$, m_i and k_i, satisfying $0 < \delta_i \leq \delta_{i-1}$, $m_i > M_3(k_{i-1})$, $k_i > k_{i-1}$, so that

$$\alpha_{m \infty k_i \infty} \text{ exists,} \qquad m \geq M_3(k_i);$$

$$\operatorname{osc}\, \alpha_{m_i n k_i \infty} > \delta_i (B_1 + 1).$$

Let $U_{kl} = 1$ for $k = k_0$, $l = 0, 1, \cdots$; $U_{k_i l} = 4^{-i} \prod_{j=0}^{i-1} \delta_j$ for $i = 1, 2, \cdots$; $U_{kl} = 0$, otherwise.

In order that the simple series, $\sum_{k=0}^{\infty} a_k U_k$, converge whenever $\sum_{k=0}^{\infty} | U_k - U_{k-1} |$ converges, the convergence of $\sum_{k=0}^{\infty} a_k$ is necessary and sufficient. Therefore $\sum_{l=0}^{\infty} a_{m_i n k l} U_{kl}$ converges, $k, i, n = 0, 1, \cdots$, as a consequence of (10). Let $S(i, j) = \sum_{l=0}^{\infty} a_{m_i n k_j l} U_{k_j l}$. Since we assume the existence of $U'_{m_i n}$, we have

$$U'_{m_i n} = \sum_{j=0}^{i-1} S(i, j) + S(i, i) + \sum_{j=i+1}^{\infty} S(i, j).$$

By examining the three parts of this expansion, we conclude, as in the proof of Theorem 12, that $\{U'_{mn}\}$ is not *urc*.

THEOREM 15. *$N \cdot a \rightarrow urc$ is* (23).

Proof. By denial of (23) we may assume the existence of an increasing sequence $\{m_i\}$ such that $\lim_{n\to\infty} A_{m_i n}$ does not exist. See proof of Theorem 8 ([3], p. 44, Theorem 20).

THEOREM 16. *$N \cdot arn \rightarrow urcrn$ is* (24).

Proof. By Theorems 4 and 12, since $ub \supset urc \supset urcrn$, we may assume (12) and (20) satisfied. By denial of (24), we can find $\{k_i\}$, $\{l_i\}$ and an increasing sequence $\{m_i\}$, $i = 0, 1, \cdots$, so that $a_{m_i \infty k_i l_i} \neq 0$.

(a) Assume $k_i = k_0$, $l_i = l_0$; let $U_{kl} = 1$ for $k = k_0$, $l = l_0$; $U_{kl} = 0$, otherwise.

(b) From the preceding paragraph it follows that there exist constants $M_5(k, l)$ such that $\lim_{n\to\infty} a_{mnkl} = 0$, $m > M_5(k, l)$, $k, l = 0, 1, \cdots$. We can choose δ_0, m_0, k_0, l_0 satisfying $0 < \delta_0 \leq (B_1 + 1)^{-1}$ so that $| a_{m_0 n k_0 l_0} | < \delta_0(B_1 + 1)$, $n = n_{0t}$, $t = 0, 1, \cdots$. We can choose for $i = 1, 2, \cdots$, δ_i, m_i, k_i and l_i satisfying $0 < \delta_i \leq \delta_{i-1}$, $m_i > m_{i-1}$, $k_i \geq k_{i-1}$, $l_i \geq l_{i-1}$ (with at most one equality sign holding in the last two conditions for a given i) so that

$$\lim_{n\to\infty} a_{m_i n k_j l_j} = 0, \qquad j = 0, 1, \cdots, i-1;$$

$$| a_{m_i n k_i l_i} | > \delta_i(B_1 + 1), \qquad n = n_{it},\ t = 0, 1, \cdots.$$

Define U_{kl} as in Theorem 12. The remainder of the proof is similar to that of Theorem 12.

THEOREM 17. *$N \cdot aurn \to urcrn$ is* (25).

Proof. By Theorem 13 we may assume (21) satisfied. By denial of (25), there exists an increasing sequence $\{m_i\}$ such that $\alpha_{m_i \infty k_0 \infty} \neq 0$. See the proof of Theorem 7.

THEOREM 18. *$N \cdot an \to urcrn$ is* (26).

Proof. See the proof of Theorem 14, assuming (25). This proof is analogous to that of Theorem 14 in the same way that the proof of Theorem 16 is analogous to that of Theorem 12.

THEOREM 19. *$N \cdot a \to urcrn$ is* (27).

Proof. By Theorem 15, since $urc \supset urcrn$, we may assume (23) satisfied. By denial of (27), there exists an increasing sequence $\{m_i\}$ such that $A_{m_i \infty} \neq 0$. See the proof of Theorem 8 ([3], p. 46, Theorem 27).

THEOREM 20. *$N \cdot arn \to rc$ is* (28).

Proof. By denial of (28), m_0, k_0 and l_0 exist such that $\lim_{n \to \infty} a_{m_0 n k_0 l_0}$ does not exist. See the proof of Theorem 6 ([3], p. 46, Theorem 31).

THEOREM 21. *$N \cdot aurn \to rc$ is* (29).

Proof. By denial of (29), m_0 and k_0 exist such that $\lim_{n \to \infty} \sum_{l=0}^{\infty} a_{m_0 n k_0 l}$ does not exist. See the proof of Theorem 7 ([3], p. 46, Theorem 32).

THEOREM 22. *$N \cdot a \to rc$ is* (30).

Proof. By denial of (30), m_0 exists such that $\lim_{n \to \infty} \sum_{k,l=0}^{\infty} a_{m_0 n k l}$ does not exist. See the proof of Theorem 8 ([3], p. 46, Theorem 33).

THEOREM 23. *$N \cdot arn \to rcrn$ is* (31).

Proof. By denial of (31), m_0, k_0 and l_0 exist such that $a_{m_0 \infty k_0 l_0} \neq 0$. See the proof of Theorem 6 ([3], p. 46, Theorem 36).

THEOREM 24. *$N \cdot aurn \to rcrn$ is* (32).

Proof. By denial of (32), m_0 and k_0 exist such that $\alpha_{m_0 \infty k_0 \infty} \neq 0$. See the proof of Theorem 7 ([3], p. 46, Theorem 37).

THEOREM 25. *$N \cdot a \to rcrn$ is* (33).

Proof. By denial of (33), m_0 exists such that $A_{m_0\infty} \neq 0$. See the proof of Theorem 8 ([3], p. 46, Theorem 38).

THEOREM 26. *$N \cdot arn \to a$ is* (34).

Proof. By Theorem 5, since $b \supset a$, we may assume (13) satisfied. Let $\sum_{k=p_i, l=q_i}^{p_i+r_i, q_i+s_i} b_{mnkl} = S(m, n, i)$. By denial of (34), either (a) $\sum_{m,n=0}^{\infty} |S(m, n, i)|$ diverges for p_i, q_i, r_i, $s_i = p_0$, q_0, r_0, s_0, respectively, or (b) there exists a sequence $\{\sum_{m,n=0}^{\infty} |S(m, n, i)|\}$, $i = 0, 1, \cdots$, which becomes infinite as i becomes infinite.

(a) Let $U_{kl} = 1$, for $k = p_0, \cdots, p_0 + r_0$ and $l = q_0, \cdots, q_0 + s_0$; $U_{kl} = 0$, otherwise. Then $\sum_{m,n=0}^{\infty} |u'_{mn}|$ diverges.

(b) Choose $\sum_{m=m_0, n=n_0}^{m_0+v_0, n_0+w_0} |S(m, n, 0)| > 1$. It follows from the preceding paragraph that there exist constants $B_2(i)$, such that

$$\sum_{m,n=0}^{\infty} \Big| \sum_{k=p, l=q}^{p+r, q+s} b_{mnkl} \Big| < B_2(i), \tag{55}$$
$$p + r \leq p_i + r_i \text{ and } q + s \leq q_i + s_i,$$
$$p = 0, 1, \cdots, p_i + r_i,\ q = 0, 1, \cdots, q_i + s_i;$$

as a consequence of (13), there exist constants, $B_1(i)$, such that

$$\sum_{m,n=0}^{m_i+v_i, n_i+w_i} \Big| \sum_{k=p, l=q}^{p+r, q+s} b_{mnkl} \Big| < 4(m_i + v_i + 1)(n_i + w_i + 1) B_1 \tag{56}$$
$$< B_1(i),\ p, q, r, s = 0, 1, \cdots.$$

Therefore, for $i = 1, 2, \cdots$, we may assume it possible to choose p_i and r_i to satisfy either $p_i = p_0$, $r_i = r_0$ or $p_i > p_{i-1} + r_{i-1}$, and m_i and v_i to satisfy either $m_i = m_0$, $v_i = v_0$ or $m_i > m_{i-1} + v_{i-1}$, and, necessarily q_i, s_i, n_i and w_i to satisfy $q_i > q_{i-1} + s_{i-1}$ and $n_i > n_{i-1} + w_{i-1}$ so that

$$\sum_{m=m_i, n=n_i}^{m_i+v_i, n_i+w_i} |S(m, n, i)| > 2B_1(i-1)[1 + 2^i + B_2(0)$$
$$+ 2^{-i} \sum_{j=1}^{i-1} B_2(j)/B_1(j-1)],$$

where $\sum_{j=1}^{0} = 0$. From this inequality and the definition of $B_1(i)$, it follows

that $B_1(i) > 2B_1(i-1)$. Let $U_{kl}=1$, $k=p_0,\cdots,p_0+r_0$ and $l=q_0,\cdots,q_0+s_0$; $U_{kl}=[2B_1(i-1)]^{-1}$, $k=p_i,\cdots,p_i+r_i$ and $l=q_i,\cdots,q_i+s_i$, $i=1,2,\cdots$; $U_{kl}=0$, otherwise.

Using (55) and (56) we have, for $i=1,2,\cdots$,

$$\begin{aligned}
\sum_{m=m_i,n=n_i}^{m_i+v_i,n_i+w_i} | u'_{mn} | &= \sum_{m=m_i,n=n_i}^{m_i+v_i,n_i+w_i} \Big| \sum_{i=0}^{\infty} S(m,n,i) U_{p_iq_i} \Big| \\
&\geq \sum_{m=m_i,n=n_i}^{m_i+v_i,n_i+w_i} \Big[| S(m,n,i) U_{p_iq_i} | - \sum_{j=0}^{i-1} | S(m,n,j) U_{p_jq_j} | \\
&\quad - \sum_{j=i+1}^{\infty} | S(m,n,j) U_{p_jq_j} | \Big] \\
&> 1 + 2^i + B_2(0) + 2^{-1} \sum_{j=1}^{i-1} B_2(j)/B_1(j-1) - \sum_{j=0}^{i-1} B_2(j) | U_{p_jq_j} | \\
&\quad - B_1(i) \sum_{j=i+1}^{\infty} | U_{p_jq_j} | \\
&> 1 + 2^i + 2^{-1}\Big\{ \sum_{j=1}^{i-1} B_2(j)/B_1(j-1) - \sum_{j=1}^{i-1} B_2(j)/B_1(j-1) \\
&\quad - B_1(i) \sum_{j=1}^{\infty} [B_1(j)]^{-1} \Big\} \\
&> 1 + 2^i - B_1(i) \sum_{j=1}^{\infty} 2^{-j} [B_1(i)]^{-1} = 2^i.
\end{aligned}$$

6. Sufficiency theorems. We shall use $S \cdot x \to y$ is (a) to indicate for sufficiency the statement corresponding to that given at the beginning of Section **5** for necessity. Let $X \supset x$ and $Y \supset y$; it is obvious that if condition (a) is $S \cdot X \to y$, it is also $S \cdot x \to y$, $S \cdot x \to Y$ and $S \cdot X \to Y$. For example, conditions which are $S \cdot an \to y$ are also $S \cdot aurn \to y$, although conditions $N \cdot an \to y$ may not be $N \cdot aurn \to y$. In each case in which the sufficient conditions for the transformation of sequences of types *aurn* and *an* into any given type are the same, the theorem for $S \cdot aurn$ is omitted.

THEOREM 27. *$S \cdot arn \to e$ is (9).*

Proof. Assume (9) satisfied. We can choose $\epsilon_1 = \epsilon[8B_1(m,n)]^{-1}$ and $P(m,n)$, $Q(m,n)$ so that for $p > P(m,n)$, $q > Q(m,n)$, we have

$$| U_{pq} | < \epsilon_1, \tag{57}$$

and

$$\sum_{i=p+1,j=0}^{p+r,q+s} | u_{ij} | + \sum_{i=0,j=q+1}^{p,q+s} | u_{ij} | < \epsilon_1, \qquad r,s=0,1,\cdots. \tag{58}$$

As a consequence of (58), we have

$$(59)\qquad \sum_{l=Q(m,n)+1}^{q} \left| \sum_{k=0}^{p+r} u_{kl} \right| < \epsilon_1; \qquad \sum_{k=P(m,n)+1}^{p} \left| \sum_{l=0}^{q+s} u_{kl} \right| < \epsilon_1;$$

and

$$(60)\qquad \sum_{l=0}^{q} \left| \sum_{k=p+1}^{p+r} u_{kl} \right| < \epsilon_1; \qquad \sum_{k=0}^{p} \left| \sum_{l=q+1}^{q+s} u_{kl} \right| < \epsilon_1.$$

We can choose $P'(m,n) > P(m,n)$, $Q'(m,n) > Q(m,n)$ so that

$$(61)\qquad \begin{aligned} &\sum_{l=0}^{Q(m,n)} \left| \sum_{k=0}^{p+r} u_{kl} \right| < \epsilon_1, && p > P'(m,n),\ r = 0, 1, \cdots, \\ &\sum_{k=0}^{P(m,n)} \left| \sum_{l=0}^{q+s} u_{kl} \right| < \epsilon_1, && q > Q'(m,n),\ s = 0, 1, \cdots. \end{aligned}$$

Let

$$S = \sum_{k,l=0}^{p+r,q+s} a_{mnkl} U_{kl} - \sum_{k,l=0}^{p-1,q-1} a_{mnkl} U_{kl};$$

$$S_1 = \sum_{l=0}^{q-1} \left[U_{pl} \sum_{k=p}^{p+r} a_{mnkl} \right];$$

$$S_2 = \sum_{k=0}^{p-1} \left[U_{kq} \sum_{l=q}^{q+s} a_{mnkl} \right];$$

$$S_3 = \sum_{k=p,\, l=q}^{p+r,q+s} a_{mnkl} U_{pq};$$

$$S_4 = \sum_{i=p+1,\, j=0}^{p+r,q+s} \sum_{k=i,\, l=j}^{p+r,q+s} a_{mnkl} u_{ij} + \sum_{i=0,\, j=q+1}^{p,q+s} \sum_{k=i,\, l=j}^{p+r,q+s} a_{mnkl} u_{ij}.$$

Then

$$\begin{aligned} S &= \sum_{k=p,\, l=0}^{p+r,q-1} a_{mnkl} U_{kl} + \sum_{k=0,\, l=q}^{p-1,q+s} a_{mnkl} U_{kl} + \sum_{k=p,\, l=q}^{p+r,q+s} a_{mnkl} U_{kl} \\ &= S_1 + \sum_{k=p+1,\, l=0}^{p+r,q-1} \sum_{i=p+1,\, j=0}^{k,l} a_{mnkl} u_{ij} \\ &\quad + S_2 + \sum_{k=0,\, l=q+1}^{p-1,q+s} \sum_{i=0,\, j=q+1}^{k,l} a_{mnkl} u_{ij} \\ &\quad + S_3 + \sum_{k=p+1,\, l=q}^{p+r,q+s} \sum_{i=p+1,\, j=0}^{k,l} a_{mnkl} u_{ij} + \sum_{k=p,\, l=q+1}^{p+r,q+s} \sum_{i=0,\, j=q+1}^{p,l} a_{mnkl} u_{ij}. \end{aligned}$$

Since

$$\begin{aligned} &\sum_{k=p+1,\, l=0}^{p+r,q-1} \sum_{i=p+1,\, j=0}^{k,l} a_{mnkl} u_{ij} + \sum_{k=p+1,\, l=q}^{p+r,q+s} \sum_{i=p+1,\, j=0}^{k,l} a_{mnkl} u_{ij} \\ &\qquad = \sum_{k=p+1,\, l=0}^{p+r,q+s} \sum_{i=p+1,\, j=0}^{k,l} a_{mnkl} u_{ij} \\ &\qquad = \sum_{i=p+1,\, j=0}^{p+r,q+s} \sum_{k=i,\, l=j}^{p+r,q+s} a_{mnkl} u_{ij}, \end{aligned}$$

and

$$\sum_{k=0,\,l=q+1}^{p-1,\,q+s}\ \sum_{i=0,\,j=q+1}^{k,\,l} a_{mnkl}u_{ij} + \sum_{k=p,\,l=q+1}^{p+r,\,q+s}\ \sum_{i=0,\,j=q+1}^{p,\,l} a_{mnkl}u_{ij}$$

$$= \sum_{i=0,\,j=q+1}^{p-1,\,q+s}\ \sum_{k=i,\,l=j}^{p-1,\,q+s} a_{mnkl}u_{ij} + \sum_{i=0,\,j=q+1}^{p,\,q+s}\ \sum_{k=p,\,l=j}^{p+r,\,q+s} a_{mnkl}u_{ij}$$

$$= \sum_{i=0,\,j=q+1}^{p,\,q+s}\ \sum_{k=i,\,l=j}^{p+r,\,q+s} a_{mnkl}u_{ij},$$

we have $S = S_1 + S_2 + S_3 + S_4$.

Using (9), (59), (60), and (61), we have for $p > P'(m, n)$, $q > Q(m, n)$, $r, s = 0, 1, \cdots,$

$$\begin{aligned} | S_1 + S_2 | &= \Big| \sum_{j=0}^{q}\ \sum_{k=p,\,l=j}^{p+r,\,q+s} a_{mnkl} \sum_{i=0}^{p} u_{ij} \Big| \\ &\leq B_1(m, n) \sum_{j=0}^{q} \Big| \sum_{i=0}^{p} u_{ij} \Big| \\ &\leq B_1(m, n) \Big[\sum_{j=0}^{Q(m,n)} \Big| \sum_{i=0}^{p+r} u_{ij} \Big| + \sum_{j=Q(m,n)+1}^{q} \Big| \sum_{i=0}^{p+r} u_{ij} \Big| + \sum_{j=0}^{q} \Big| \sum_{i=p+1}^{p+r} u_{ij} \Big| \\ &< 3(\epsilon/8). \end{aligned}$$

Similarly we can show that $| S_2 + S_3 | < 3\epsilon/8$, $p > P(m, n)$, $q > Q'(m, n)$, $r, s = 0, 1, \cdots$.

As a consequence of (9) and (57), for $p > P(m, n)$, $q > Q(m, n)$, we have $| S_3 | = \big| \sum_{k=p,\,l=q}^{p+r,\,q+s} a_{mnkl} \big| \cdot | U_{pq} | < B_1(m, n) \cdot \epsilon_1 = \epsilon/8$. As a consequence of (9) and (58), for $p > P(m, n)$, $q > Q(m, n)$, we have

$$\begin{aligned} | S_4 | &\leq \sum_{i=p+1,\,j=0}^{p+r,\,q+s} \Big| \sum_{k=i,\,l=j}^{p+r,\,q+s} a_{mnkl} \Big|\ | u_{ij} | + \sum_{i=0,\,j=q+1}^{p,\,q+s} \Big| \sum_{k=i,\,l=j}^{p+r,\,q+s} a_{mnkl} \Big|\ | u_{ij} | \\ &< B_1(m, n) \Big[\sum_{i=p+1,\,j=0}^{p+r,\,q+s} | u_{ij} | + \sum_{i=0,\,j=q+1}^{p,\,q+s} | u_{ij} | \\ &< B_1(m, n) \cdot \epsilon_1 = \epsilon/8. \end{aligned}$$

Therefore for $p > P'(m, n)$, $q > Q'(m, n)$, $r, s = 0, 1, \cdots,$

$$\begin{aligned} | S | &= | (S_1 + S_3) + (S_2 + S_3) - S_3 + S_4 | \\ &\leq | S_1 + S_3 | + | S_2 + S_3 | + | S_3 | + | S_4 | < \epsilon, \end{aligned}$$

and U'_{mn} exists, $m, n = 0, 1, \cdots$.

THEOREM 28. *$S \cdot an \to e$ are* (9) *and* (10).

Proof. Assume (9) and (10) satisfied. We can choose $\epsilon_1 = \epsilon[8B_1(m,n)]^{-1}$ and $P(m,n)$, $Q(m,n)$, so that for $p > P(m,n)$, $q > Q(m,n)$, (57) and (58) are satisfied; we can choose $\epsilon_2 = \epsilon[8(Q+1)U^*]^{-1}$ and $P'(m,n) > P(m,n)$, $Q'(m,n) > Q(m,n)$ so that

$$(62)\qquad \begin{aligned} &\Big|\sum_{k=p}^{p+r} a_{mnkl}\Big| < \epsilon_2, \qquad p > P'(m,n)\,;\ r = 0,1,\cdots;\ l = 0,1,\cdots,Q,\\ &\Big|\sum_{l=q}^{q+s} a_{mnkl}\Big| < \epsilon_2, \qquad q > Q'(m,n)\,;\ s = 0,1,\cdots;\ k = 0,1,\cdots,P. \end{aligned}$$

Assume S, S_1, S_2, S_3 and S_4 defined as in Theorem 27; let

$$\begin{aligned} T_1 &= \sum_{j=0}^{Q}\Big[\sum_{i=0}^{P} u_{ij} \sum_{k=p,\,l=j}^{p+r,\,Q} a_{mnkl}\Big]\\ &= \sum_{l=0}^{Q}\Big[\sum_{k=p}^{p+r} a_{mnkl} \sum_{i,j=0}^{p,\,l} u_{ij}\Big]; \end{aligned}$$

$$T_2 = U_{p,Q+1} \sum_{k=p,\,l=Q+1}^{p+r,\,q-1} a_{mnkl};$$

$$T_3 = \sum_{j=Q+2}^{q-1}\Big[\sum_{i=0}^{p} u_{ij} \sum_{k=p,\,l=j}^{p+r,\,q-1} a_{mnkl}\Big].$$

Then since $T_1 = \sum_{j=0}^{Q} (U_{pj} - U_{p,j-1}) \sum_{k=p,\,l=j}^{p+r,\,Q} a_{mnkl} = \sum_{l=0}^{Q} U_{pl} \sum_{k=p}^{p+r} a_{mnkl}$, and since

$$\begin{aligned} T_2 + T_3 &= U_{p,Q+1} \sum_{k=p,\,l=Q+1}^{p+r,\,q-1} a_{mnkl} + \sum_{j=Q+2}^{q-1} (U_{pj} - U_{p,j-1}) \sum_{k=p,\,l=j}^{p+r,\,q-1} a_{mnkl}\\ &= \sum_{l=Q+1}^{q-1} U_{pl} \sum_{k=p}^{p+r} a_{mnkl}, \end{aligned}$$

we have $T_1 + T_2 + T_3 = S_1$. Using (62), we have $|T_1| < \epsilon/8$, $p > P'(m,n)$; using (57) and (9), we have $|T_2| < \epsilon/8$, $p > P(m,n)$, $q > Q(m,n)$; using (58) and (9), we have $|T_3| < \epsilon/8$, $p > P(m,n)$, $q > Q(m,n)$. Therefore $|S_1| = |T_1 + T_2 + T_3| < 3\epsilon/8$, $p > P'(m,n)$, $q > Q(m,n)$. Similarly we can show that $|S_2| < 3\epsilon/8$, $p > P(m,n)$, $q > Q'(m,n)$. It was proved in Theorem 27 that $|S_3| < \epsilon/8$, $|S_4| < \epsilon/8$, $p > P(m,n)$, $q > Q(m,n)$. Therefore for $p > P'(m,n)$, $q > Q'(m,n)$, $r, s = 0, 1, \cdots$,

$$|S| \leq |S_1| + |S_2| + |S_3| + |S_4| < \epsilon.$$

THEOREM 29. *$S \cdot a \to e$ are* (10) *and* (11).

Proof. Using (35) and Theorem 28, we know that since $\{U_{kl} - U\}$ is *an*, $\sum_{k,l=0}^{\infty} a_{mnkl}(U_{kl} - U)$ converges. Using (11), we have $\sum_{k,l=0}^{\infty} a_{mnkl}U = A_{mn}U$. Therefore U'_{mn} exists, since

$$\sum_{k,l=0}^{\infty} a_{mnkl}(U_{kl} - U) + \sum_{k,l=0}^{\infty} a_{mnkl}U = U'_{mn}.$$

THEOREM 30. *$S \cdot arn \to ub$ is* (12).

Proof. Since (12) $\to$ (9), U'_{mn} exists; we have, for $m, n \geq M_1$,

$$\begin{aligned} |U'_{mn}| &= \lim_{p,q\to\infty} \Big|\sum_{k,l=0}^{p,q} a_{mnkl}U_{kl}\Big| \\ &= \lim_{p,q\to\infty} \Big|\sum_{i,j=0}^{p,q} \sum_{k=i,\,l=j}^{p,q} a_{mnkl}u_{ij}\Big| \\ &< B_1 \lim_{p,q\to\infty} \sum_{i,j=0}^{p,q} |u_{ij}| = B_1U^*. \end{aligned}$$

THEOREM 31. *$S \cdot an \to ub$ are* (10) *and* (12).

THEOREM 32. *$S \cdot a \to ub$ are* (10), (11) *and* (12).

THEOREM 33. *$S \cdot arn \to b$ is* (13).

THEOREM 34. *$S \cdot an \to b$ are* (10) *and* (13).

THEOREM 35. *$S \cdot a \to b$ are* (10), (11), *and* (13).

THEOREM 36. *$S \cdot arn \to c$ are* (12) *and* (14).

Proof. Let

$$T_1 = \sum_{k,l=0}^{p-1,q-1} a_{mnkl}U_{kl};$$

$$T_2 = \sum_{k,l=0}^{p-1,q-1} a_{kl}U_{kl};$$

$$T_3 = \sum_{k,l=0}^{\infty} a_{kl}U_{kl}.$$

Using the proof of Theorem 27, with (9) replaced by (12), we find constants P' and Q' such that $\Big|\sum_{k,l=0}^{p+r,q+s} a_{mnkl}U_{kl} - T_1\Big| < \epsilon/3$, for $p > P'$, $q > Q'$,

$r, s = 0, 1, \cdots, m, n \geqq M_1$; that is, $| U'_{mn} - T_1 | < \epsilon/3$, $p > P'$, $q > Q'$, $m, n \geqq M_1$. As a consequence of (37), $\sum_{k,l=0}^{\infty} a_{kl} U_{kl}$ converges and we can choose P'', Q'' satisfying $P'' > P'$, $Q'' > Q'$, so that $| T_2 - T_3 | < \epsilon/3$, $p > P''$, $q > Q''$. We choose arbitrary integers P and Q satisfying $P > P''$, $Q > Q''$; we can choose $M > M_1$ so that for $m, n \geqq M$, $k = 0, 1, \cdots, P-1$, $l = 0, 1, \cdots, Q-1$,

$$| a_{mnkl} - a_{kl} | < \epsilon [3PQU']^{-1}.$$

It follows that for $p = P$, $q = Q$, $m, n \geqq M$, $| T_1 - T_2 | < \epsilon/3$. Therefore for $m, n > M$,

$$| U'_{mn} - T_3 | = | (U'_{mn} - T_1) + (T_1 - T_2) + (T_2 - T_3) | < \epsilon,$$

and

$$U' = \sum_{k,l=0}^{\infty} a_{kl} U_{kl}.$$

THEOREM 37. *$S \cdot an \to c$ are* (12), (14) *and* (15).

Proof. The double sequence $\{U_{kl} - U_{k\infty} - U_{\infty l}\}$ is *arn*; since (12) and (14) are satisfied,

$$\lim_{m,n\to\infty} \sum_{k,l=0}^{\infty} a_{mnkl} (U_{kl} - U_{k\infty} - U_{\infty l}) = \sum_{k,l=0}^{\infty} a_{kl} (U_{kl} - U_{k\infty} - U_{\infty l}).$$

The double sequence $\{U_{k\infty}\}$ is *an*. Since (9) and (10) are satisfied, $\sum_{k,l=0}^{\infty} a_{mnkl} U_{k\infty}$ exists, $m, n = 0, 1, \cdots$, and $\sum_{k=0}^{\infty} a_{mnkl} U_{k\infty} = \alpha_{mnk\infty} U_{k\infty}$; therefore $\sum_{k=0}^{\infty} \alpha_{mnk\infty} U_{k\infty} = \sum_{k,l=0}^{\infty} a_{mnkl} U_{k\infty}$. It follows that we can find P′ so that

$$(63) \qquad \left| \sum_{k,l=0}^{\infty} a_{mnkl} U_{k\infty} - \sum_{k=0}^{p-1} \alpha_{mnk\infty} U_{k\infty} \right| < \epsilon/3, \quad p > P'.$$

In order that the simple series $\sum_{k=0}^{\infty} a_k U_k$ converge whenever $\sum_{k=0}^{\infty} | U_k - U_{k-1} |$ converges and $\lim_{k\to\infty} U_k = 0$, it is necessary and sufficient that there exist a constant B_1 such that $| \sum_{k=0}^{p} a_k | < B_1$, $p = 0, 1, \cdots$; therefore, as a consequence of (38) we can find P'' satisfying $P'' > P'$, so that

$$\left| \sum_{k=0}^{p-1} \alpha_{k\infty} U_{k\infty} - \sum_{k=0}^{\infty} \alpha_{k\infty} U_{k\infty} \right| < \epsilon/3, \qquad p > P''.$$

Since (15) is satisfied, for an arbitrary P satisfying $P > P''$, we can find M so that, for $p = P$,

$$\left| \sum_{k=0}^{p-1} \alpha_{mnk\infty} U_{k\infty} - \sum_{k=0}^{p-1} \alpha_{k\infty} U_{k\infty} \right| < \epsilon/3, \qquad m, n \geq M.$$

Therefore for $m, n \geq M$,

$$\left| \sum_{k,l=0}^{\infty} a_{mnkl} U_{k\infty} - \sum_{k=0}^{\infty} \alpha_{k\infty} U_{k\infty} \right| < \epsilon.$$

It follows that

$$(64) \qquad U' = \lim_{m,n\to\infty} \sum_{k,l=0}^{\infty} a_{mnkl}[(U_{kl} - U_{k\infty} - U_{\infty l}) + U_{k\infty} + U_{\infty l}]$$

$$= \sum_{k,l=0}^{\infty} a_{kl}(U_{kl} - U_{k\infty} - U_{\infty l}) + \sum_{k=0}^{\infty} \alpha_{k\infty} U_{k\infty} + \sum_{l=0}^{\infty} \alpha_{\infty l} U_{\infty l}.$$

THEOREM 38. *$S \cdot a \to c$ are* (12), (14), (15) *and* (16).

Proof. Since the double sequences $\{U\}$ and $\{U_{kl} - U\}$ are a and an respectively, $\sum_{k,l=0}^{\infty} a_{mnkl} U$ and $\sum_{k,l=0}^{\infty} a_{mnkl}(U_{kl} - U)$ exist, $m, n = 0, 1, \cdots$. Since (16) is satisfied,

$$\lim_{m,n\to\infty} \sum_{k,l=0}^{\infty} a_{mnkl} U = UA;$$

since the conditions of Theorem 37 are satisfied,

$$(65) \qquad U' = \lim_{m,n\to\infty} \sum_{k,l=0}^{\infty} a_{mnkl}[U + (U_{kl} - U)]$$

$$= UA + \sum_{k,l=0}^{\infty} a_{kl}(U_{kl} + U - U_{k\infty} - U_{\infty l}) + \sum_{k=0}^{\infty} \alpha_{k\infty}(U_{k\infty} - U)$$

$$+ \sum_{l=0}^{\infty} \alpha_{\infty l}(U_{\infty l} - U).$$

THEOREM 39. *$S \cdot arn \to bc$ are* (13) *and* (14).

THEOREM 40. *$S \cdot an \to bc$ are* (13), (14) *and* (15).

THEOREM 41. *$S \cdot a \to bc$ are* (13), (14), (15) *and* (16).

THEOREM 42. *$S \cdot arn \to cn$ are* (12) *and* (17).

THEOREM 43. *$S \cdot an \to cn$ are* (12), (17) *and* (18).

THEOREM 44. *$S \cdot a \to cn$ are (12), (17), (18) and (19).*

THEOREM 45. *$S \cdot arn \to bcn$ are (13) and (17).*

THEOREM 46. *$S \cdot an \to bcn$ are (13), (17) and (18).*

THEOREM 47. *$S \cdot a \to bcn$ are (13), (17), (18) and (19).*

THEOREM 48. *$S \cdot arn \to urc$ are (12), (14) and (20).*

Proof. Let T_1, T_2 and T_3 be equal respectively to T_1, T_2 and T_3 of Theorem 36, with a_{kl} replaced by $a_{m\infty kl}$. Let m represent an arbitrary constant satisfying $m \geq M_1 + M_2$. Using (20) and replacing (37) by (39), we prove (as in Theorem 36) the existence of constants P, Q and N such that the following relationships are satisfied: $| U'_{mn} - T_1 | < \epsilon/3$ and $| T_1 - T_2 | < \epsilon/3$, $p = P$, $q = Q$, $n \geq N$; $\sum_{k,l=0}^{\infty} a_{m\infty kl} U_{kl}$ exists and $| T_2 - T_3 | < \epsilon/3$, $p = P$, $q = Q$. Therefore

$$U'_{m\infty} = \sum_{k,l=0}^{\infty} a_{m\infty kl} U_{kl}, \qquad m \geq M_1 + M_2. \tag{66}$$

THEOREM 49. *$S \cdot aurn \to urc$ are (12), (14), (15), (20) and (21).*

Proof. As a consequence of Theorem 48, for $m \geq M_1 + M_2$,

$$\lim_{n\to\infty} \sum_{k,l=0}^{\infty} a_{mnkl}(U_{kl} - U_{k\infty} - U_{\infty l}) = \sum_{k,l=0}^{\infty} a_{m\infty kl}(U_{kl} - U_{k\infty} - U_{\infty l}). \tag{67}$$

We can find P' as in Theorem 37, so that (63) is satisfied. Since $\{U_{kl}\}$ is *aurn*, we can find P'' satisfying $P'' > P'$ so that $U_{k\infty} = 0$, $k > P''$; therefore we have

$$| \sum_{k=0}^{p-1} \alpha_{m\infty k\infty} U_{k\infty} - \sum_{k=0}^{\infty} \alpha_{m\infty k\infty} U_{k\infty} | < \epsilon/3, \qquad p > P''.$$

Since (21) is satisfied, for an arbitrary p satisfying $p > P''$, we can find $M > M_1 + M_2$ so that

$$| \sum_{k=0}^{p-1} \alpha_{mnk\infty} U_{k\infty} - \sum_{k=0}^{p-1} \alpha_{m\infty k\infty} U_{k\infty} | < \epsilon/3, \qquad m \geq M.$$

Therefore for $m \geq M$,

$$| \sum_{k,l=0}^{\infty} a_{mnkl} U_{k\infty} - \sum_{k=0}^{\infty} \alpha_{m\infty k\infty} U_{k\infty} | < \epsilon.$$

It follows that for $m \geq M$,

$$(68) \qquad U'_{m\infty} = \lim_{n\to\infty} \sum_{k,l=0}^{\infty} a_{mnkl}[(U_{kl} - U_{k\infty} - U_{\infty l}) + U_{k\infty} + U_{\infty l}]$$

$$= \sum_{k,l=0}^{\infty} a_{m\infty kl}(U_{kl} - U_{k\infty} - U_{\infty l}) + \sum_{k=0}^{\infty} \alpha_{m\infty k\infty} U_{k\infty}$$

$$+ \sum_{l=0}^{\infty} \alpha_{m\infty\infty l} U_{\infty l}.$$

THEOREM 50. *$S \cdot an \to urc$ are* (12), (14), (15), (20) *and* (22).

Proof. The double sequence $\{U_{kl} - U_{k\infty} - U_{\infty l}\}$ is *arn*; therefore (67) is satisfied. The rest of the proof is the same as that of Theorem 37, with $\alpha_{k\infty}$, (38), (15) and (64) replaced by $\alpha_{m\infty k\infty}$, (40), (22) and (68) respectively.

THEOREM 51. *$S \cdot a \to urc$ are* (12), (14), (15), (16), (20), (22) *and* (23).

Proof. Since (23) is satisfied, $\sum_{k,l=0}^{\infty} a_{mnkl}U$ exists, $m, n = 0, 1, \cdots$, and approaches $UA_{m\infty}$ as $n \to \infty$, $m \geq M_4$. Since the conditions of Theorem 50 are satisfied, $\sum_{k,l=0}^{\infty} a_{mnkl}(U_{kl} - U)$ exists, $m, n = 0, 1, \cdots$, and approaches a limit as $n \to \infty$, $m \geq M_1 + M_2 + M_3$. Therefore, for $m \geq M_1 + M_2 + M_3 + M_4$,

$$(69) \qquad U'_{m\infty} = UA_{m\infty} + \sum_{k,l=0}^{\infty} a_{m\infty kl}(U_{kl} + U - U_{k\infty} - U_{\infty l})$$

$$+ \sum_{k=0}^{\infty} \alpha_{m\infty k\infty}(U_{k\infty} - U) + \sum_{l=0}^{\infty} \alpha_{m\infty\infty l}(U_{\infty l} - U).$$

THEOREM 52. *$S \cdot arn \to burc$ are* (13), (14) *and* (20).

THEOREM 53. *$S \cdot aurn \to burc$ are* (13), (14), (15), (20) *and* (21).

THEOREM 54. *$S \cdot an \to burc$ are* (13), (14), (15), (20) *and* (22).

THEOREM 55. *$S \cdot a \to burc$ are* (13), (14), (15), (16), (20), (22) *and* (23).

THEOREM 56. *$S \cdot arn \to urcn$ are* (12), (17) *and* (20).

THEOREM 57. *$S \cdot aurn \to urcn$ are* (12), (17), (18), (20) *and* (21).

THEOREM 58. *$S \cdot an \to urcn$ are* (12), (17), (18), (20) *and* (22).

THEOREM 59. *$S \cdot a \to urcn$ are* (12), (17), (18), (19), (20), (22) *and* (23).

THEOREM 60. *$S \cdot arn \to burcn$ are* (13), (17) *and* (20).

THEOREM 61. *$S \cdot aurn \to burcn$ are* (13), (17), (18), (20) *and* (21).

THEOREM 62. *$S \cdot an \to burcn$ are* (13), (17), (18), (20) *and* (22).

THEOREM 63. *$S \cdot a \to burcn$ are* (13), (17), (18), (19), (20), (22) *and* (23).

THEOREM 64. *$S \cdot arn \to urcrn$ are* (12), (14) *and* (24).

Proof. Since the conditions of Theorem 48 are satisfied, we have (66); using (24), we have $U'_{m\infty} = 0,\ m \geqq M_1 + M_2 + M_5$.

THEOREM 65. *$S \cdot aurn \to urcrn$ are* (12), (14), (15), (24) *and* (25).

Proof. Since the conditions of Theorem 49 are satisfied, we have (68). Choose k_0 and l_0 so that $U_{k\infty} = 0$, $k > k_0$, and $U_{\infty l} = 0, l > l_0$. We can choose M'_6 satisfying $M'_6 > M_6(k)$, $k = 0, 1, \cdots, k_0$ and $M'_6 > M_6(l)$, $l = 0, 1, \cdots, l_0$. Then for $m \geqq M + M_5 + M'_6$, $U'_{m\infty} = 0$.

THEOREM 66. *$S \cdot an \to urcrn$ are* (12), (14), (15), (24) *and* (26).

Proof. The conditions of Theorem 50 are satisfied; choose $m \geqq M + M_5 + M_6$.

THEOREM 67. *$S \cdot a \to urcrn$ are* (12), (14), (15), (16), (24), (26) *and* (27).

Proof. Since the conditions of Theorem 51 are satisfied, we have (69); choose $m \geqq \sum_{i=1}^{7} M_i$.

THEOREM 68. *$S \cdot arn \to burcrn$ are* (13), (14) *and* (24).

THEOREM 69. *$S \cdot aurn \to burcrn$ are* (13), (14), (15), (24) *and* (25).

THEOREM 70. *$S \cdot an \to burcrn$ are* (13), (14), (15), (24) *and* (26).

THEOREM 71. *$S \cdot a \to burcrn$ are* (13), (14), (15), (16), (24), (26) *and* (27).

THEOREM 72. *$S \cdot arn \to rc$ are* (13), (14) *and* (28).

Proof. Use the proof of Theorem 48 with $M_1 = M_2 = 0$.

THEOREM 73. *$S \cdot an \rightarrow rc$ are* (13), (14), (15), (28) *and* (29).

Proof. Use the proof of Theorem 50, with $M_1 = M_2 = M_3 = 0$.

THEOREM 74. *$S \cdot a \rightarrow rc$ are* (13), (14), (15), (16), (28), (29) *and* (30).

Proof. Use the proof of Theorem 51, with $M_i = 0$, $i = 1, 2, 3, 4$.

THEOREM 75. *$S \cdot arn \rightarrow rcn$ are* (13), (17) *and* (28).

THEOREM 76. *$S \cdot an \rightarrow rcn$ are* (13), (17), (18), (28) *and* (29).

THEOREM 77. *$S \cdot a \rightarrow rcn$ are* (13), (17), (18), (19), (28), (29) *and* (30).

THEOREM 78. *$S \cdot arn \rightarrow rcurn$ are* (13), (14), (24) *and* (28).

THEOREM 79. *$S \cdot aurn \rightarrow rcurn$ are* (13), (14), (15), (24), (25), (28) *and* (29).

THEOREM 80. *$S \cdot an \rightarrow rcurn$ are* (13), (14), (15), (24), (26), (28) *and* (29).

THEOREM 81. *$S \cdot a \rightarrow rcurn$ are* (13), (14), (15), (16), (24), (26), (27), (28), (29) *and* (30).

THEOREM 82. *$S \cdot arn \rightarrow rcrn$ are* (13), (14) *and* (31).

Proof. Use the proof of Theorem 72, with $a_{m\infty kl} = 0$.

THEOREM 83. *$S \cdot an \rightarrow rcrn$ are* (13), (14), (15), (31) *and* (32).

Proof. Use the proof of Theorem 73, with $a_{m\infty kl} = \alpha_{m\infty k\infty} = \alpha_{m\infty\infty l} = 0$.

THEOREM 84. *$S \cdot a \rightarrow rcrn$ are* (13), (14), (15), (16), (31), (32) *and* (33).

Proof. Use the proof of Theorem 74 with $a_{m\infty kl} = \alpha_{m\infty k\infty} = \alpha_{m\infty\infty l} = A_{m\infty} = 0$.

THEOREM 85. *$S \cdot arn \rightarrow a$ is* (34).

Proof. Since (34) $\rightarrow$ (9), $u'_{mn} = \sum_{k,l=0}^{\infty} b_{mnkl} U_{kl}$ exists; it follows that

for arbitrarily chosen m_0 and n_0, and for a given ϵ, we can find k_0 and l_0, dependent upon m_0, n_0 and ϵ, so that

$$| u'_{mn} - \sum_{k,l=0}^{k_0,l_0} b_{mnkl} U_{kl} | < \epsilon (m_0 + 1)^{-1} (n_0 + 1)^{-1},$$
$$m = 0, 1, \cdots, m_0; \; n = 0, 1, \cdots, n_0.$$

Therefore

$$\sum_{m,n=0}^{m_0,n_0} | u'_{mn} | < \sum_{m,n=0}^{m_0,n_0} | \sum_{k,l=0}^{k_0,l_0} b_{mnkl} U_{kl} | + \epsilon.$$

Since

$$\sum_{m,n=0}^{m_0,n_0} | \sum_{k,l=0}^{k_0,l_0} b_{mnkl} U_{kl} | = \sum_{m,n=0}^{m_0,n_0} | \sum_{p,q=0}^{k_0,l_0} \sum_{k=p,l=q}^{k_0,l_0} b_{mnkl} u_{pq} |$$
$$\leq \sum_{p,q=0}^{k_0,l_0} \sum_{m,n=0}^{m_0,n_0} | \sum_{k=p,l=q}^{k_0,l_0} b_{mnkl} | \; | u_{pq} |,$$

we have, as a consequence of (34),

$$\sum_{m,n=0}^{m_0,n_0} | u'_{mn} | < B_2 \sum_{p,q=0}^{k_0,l_0} | u_{pq} | + \epsilon$$
$$\leq B_2 U^* + \epsilon.$$

Therefore

$$\sum_{m,n=0}^{m_0,n_0} | u'_{mn} | \leq B_2 U^*$$

or

$$\sum_{m,n=0}^{\infty} | u'_{mn} | \leq B_2 U^*.$$

The proof of each of the following two theorems is the same, except for the proof of the existence of the transform, as that of Theorem 85.

THEOREM 86. *$S \cdot an \to a$ are* (10) *and* (34).

THEOREM 87. *$S \cdot a \to a$ are* (10), (11), *and* (34).

THEOREM 88. *$S \cdot arn \to an$ are* (17) *and* (34).

THEOREM 89. *$S \cdot an \to an$ are* (17), (18) *and* (34).

THEOREM 90. *$S \cdot a \to an$ are* (17), (18), (19) *and* (34).

THEOREM 91. *$S \cdot arn \to aurn$ are* (24) *and* (34).

THEOREM 92. *$S \cdot aurn \to aurn$ are* (10), (24), (25) *and* (34).

THEOREM 93. *$S \cdot an \to aurn$ are* (10), (24), (26) *and* (34).

THEOREM 94. *$S \cdot a \to aurn$ are* (10), (11), (24), (26), (27) *and* (34).

THEOREM 95. *$S \cdot arn \to arn$ are* (31) *and* (34).

THEOREM 96. *$S \cdot an \to arn$ are* (10), (31), (32) *and* (34).

THEOREM 97. *$S \cdot a \to arn$ are* (10), (11), (31), (32), (33) *and* (34).

7. Conclusion. The conditions which have been proved sufficient in Section **6** are also necessary. This can be proved by the use of the theorems of Section **5**.

If the conditions of Theorems 38, 41, 51, 55, 74 or 87 are satisfied, $a \to c$, bc, urc, $burc$, rc or a, respectively; in each case U' is found in (65). In order that $U' = UA$, it is sufficient that $a_{kl} = \alpha_{k\infty} = \alpha_{\infty l} = 0$; therefore $U' = UA$, and hence $U' = 0$ when $U = 0$, if in each of these theorems we replace conditions (14) and (15) (which are included by implication in the case of Theorem 87) by (17) and (18). If, in addition, $A = 1$, we have $U' = U$. In both cases the conditions are not only sufficient, but necessary

THE GEORGE WASHINGTON UNIVERSITY.

BIBLIOGRAPHY.

[1] Hahn, H., "Über Folgen linearer Operationen," *Monatshefte für Mathematik und Physik*, vol. 32 (1922), pp. 3-88.

[2] Hamilton, H. J., "On transformations of double series," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 275-283.

[3] ———, "Transformations of multiple sequences," *Duke Mathematical Journal*, vol. 2 (1936), pp. 29-60.

[4] Mears, F. M., "Absolute regularity and the Nörlund mean," *Annals of Mathematics*, vol. 38, no. 3 (1937), pp. 594-601.

ON THE LOCATION OF SPECTRA OF DIFFERENTIAL EQUATIONS.*

By Sylvan Wallach.

1. Let $f(t)$ be a real-valued continuous function on the half-line $0 \leqq t < \infty$, and let λ be a real parameter. The function $f(t)$ and the differential equation

$$x'' + (\lambda + f(t))x = 0 \tag{1}$$

are said to be of Type I if not every solution of (1) is of class $L^2(0, \infty)$ for some λ, (hence for every λ, cf. [8], p. 238). Otherwise $f(t)$ and (1) are said to be of Type II. (The two types correspond to the *Grenzpunktfall* and *Grenzkreisfall* of Weyl [8].) If $f(t)$ is of Type I, then it is known that (1) and the boundary conditions

$$x(0) \sin \alpha + x'(0) \cos \alpha = 0, \tag{2}$$

$$\int_0^\infty x^2(t)\, dt < \infty, \tag{3}$$

define a self-adjoint boundary value problem and determine a spectrum S_α in accordance with the theory of Hilbert's space.

If $f(t)$ is of Type I, let P_α denote the point spectrum, C_α the continuous spectrum and D_α the cluster spectrum of (1), (2) and (3), so that $S_\alpha = P_\alpha + C_\alpha + D_\alpha$. Then λ is in S_α if and only if the differential equation

$$x'' + (\lambda + f(t))x = g(t)$$

and boundary conditions (2) and (3) fail to have a solution for some continuous function $g(t)$ of class $L^2(0, \infty)$; whereas λ is in P_α if and only if (1), (2) and (3) have a solution $x(t) \not\equiv 0$; finally D_α consists of the cluster points of P_α. It is further known that $C_\alpha + D_\alpha$, the derived set of S_α, is independent of α; cf. [8], p. 251. This invariant λ-set will be called the essential spectrum and can be denoted S'.

2. We shall prove:

(I) *If*

$$\int^\infty t f^2(t)\, dt < \infty, \tag{4}$$

* Received April 22, 1948.

then

(i) (1) *is of* Type I,

(ii) *no positive* λ *is in* P_a,

(iii) *every non-negative* λ *is in* C_a,

(iv) *no negative* λ *is in* C_a,

(v) D_a *consists at most of* $\lambda = 0$,

(vi) P_a *is a bounded set* (*for fixed* α).

(II) *If*

$$\limsup_{t\to\infty} t\,|f(t)| < \infty, \tag{5}$$

then the assertions (i), (iv) *and* (vi) *of* (I) *are true. Further, if* b *denotes the limit superior in* (5), *then*

(ii*) *no* $\lambda > b^2$ *is in* P_a,

(iii*) *every* $\lambda \geqq b^2$ *is in* C_a.

The following theorem is stated for the sake of completeness.

(III) *If either*

$$\int^{\infty} |f(t)|\,dt < \infty \tag{a}$$

or

$$\int^{\infty} |df(t)| < \infty, \text{ with } f(\infty) = 0, \tag{b}$$

then (i) *to* (vi) *of* (I) *hold.*

Case (a) of (III) (with the additional assumption $f(t) = o(1)$, as $t \to \infty$) goes back to Weyl [8] (see also Titchmarsh [4]). The method of treatment depends on the determination of the spectral resolution; however, both cases in (III) are easily proved by virtue of known asymptotic formulas and oscillation theorems, without the explicit determination of spectral resolutions.

3. The assumptions of the theorems imply that $f(t)$ is small in some sense as $t \to \infty$. In (I) and in case (a) of (III) smallness is imposed by an integrability condition, so that $f(t)$ need not be bounded, whereas in the remaining theorems $f(t) = o(1)$ as $t \to \infty$. Although the condition

$f(t) = o(1)$ alone is sufficient that (i) and (iv) hold, it remains undecided whether the same is true of (iii).

4. Nevertheless, it can be shown that the theorems are the best possible in so far as assertions (ii) and (ii*) are concerned. In what follows let $\epsilon > 0$.

(I*) *Assertion* (ii) *of* (I) *is not true if* (4) *is relaxed to*

$$\int^{\infty} t^{1-\epsilon} f^2(t)\,dt < \infty. \tag{6}$$

(II*) *Assertion* (ii*) *of* (II) *is not true if the lower bound* b^2 *for* λ *is replaced by* $b^2 - \epsilon$. *Moreover* b^2 *cannot be replaced by an absolute constant.*

(III*) *Neither*

$$\int^{\infty} |f(t)|^{1+\epsilon} dt < \infty \tag{a*}$$

nor

$$\int^{\infty} |f'(t)|^{1+\epsilon} dt < \infty, \text{ with } f(\infty) = 0 \tag{b*}$$

is sufficient that (ii) *hold.*

5. Before proceeding with the proof of (I), some facts implied by a weaker assumption than (4), namely,

$$\int^{\infty} f^2(t)\,dt < \infty, \tag{7}$$

will be stated as a lemma:

(IV) *If* $f(t)$ *is of class* $L^2(0, \infty)$, *then* (1) *is of* Type I. *Moreover, if* (1) *has a solution* $x(t) \not\equiv 0$ *of class* $L^2(0, \infty)$ *for* $\lambda = \lambda_0$ *then* $x'(t)$ *and* $x''(t)$ *are of class* $L^2(0, \infty)$; $x(t)$ *and* $x'(t)$ *are* $o(1)$ *as* $t \to \infty$; *and, if* $\lambda_0 \neq 0$,

$$\lim_{t\to\infty} \int^{t} x(s)\,ds \tag{8}$$

exists; (i. e., $x(t)$ *is improperly integrable.*)

Since (7) and the assertions of (IV) are invariant under the substitution $\tau = |\lambda|^{\frac{1}{2}} t$, which replaces λ in (1) by $\operatorname{sgn} \lambda$, it is sufficient to consider only the three cases $\lambda = 0$, $\lambda = \pm 1$. Suppose then, that for $\lambda = \lambda_0 = 1$, $x(t)$ is a solution of (1) which is of class $L^2(0, \infty)$. In view of (7), $f(t)x(t)$ is of class $L(0, \infty)$, so that $x(t)$ is a solution of the integral equation

$$x(t) = \{c_1 + \int_0^t f(s)x(s) \cos s\, ds\} \sin t + \{c_2 + \int_t^\infty f(s)x(s) \sin s\, ds\} \cos t, \tag{9}$$

for a suitable choice of the constants c_1, c_2. Thus

$$x(t) = c_3 \sin t + c_2 \cos t + o(1) \quad \text{as} \quad t \to \infty, \tag{10}$$

where

$$c_3 = c_1 + \int_0^\infty f(s)x(s) \cos s\, ds.$$

In order that x be of class $L^2(0, \infty)$, it is necessary that $c_2 = c_3 = 0$ in (10). Hence (9) can be written in the form

$$x(t) = \int_t^\infty f(s)x(s) \sin (t - s) ds. \tag{11}$$

Moreover, since

$$\int^\infty | f(s)x(s) | \, ds < \infty,$$

formal differentiation of (11) is justified, so that

$$x'(t) = \int_t^\infty f(s)x(s) \cos (t - s) ds. \tag{12}$$

Formulae (11) and (12) show that x and x' are $o(1)$ as $t \to \infty$. The boundedness of x, along with (1) and (7), then implies that x'' is of class $L^2(0, \infty)$. Hence x' is also of class $L^2(0, \infty)$.

For the case $\lambda = 1$ of (1), let $y(t)$ be any solution of class $L^2(0, \infty)$. Then the Wronskian, $xy' - y'x$, is $o(1)$ as $t \to \infty$. Consequently x and y are linearly dependent, and this implies that the differential equation (1) is of Type I. The remaining assertion of (IV), for the case $\lambda_0 = 1$, follows by a quadrature of (1):

$$x'(t_2) - x'(t_1) + \int_{t_1}^{t_2} x(s) ds + \int_{t_1}^{t_2} f(s)x(s) ds = 0.$$

Since fx is of class $L(0, \infty)$, and $x' = o(1)$, the preceding formula line implies the existence of the limit (8).

The truth of (IV) for the case $\lambda_0 = -1$ can be readily inferred from known asymptotic formulae [2]. For in this case the differential equation (1) has a solution $x(t)$ of class $L^2(0, \infty)$ which satisfies

$$x(t) \sim e^{-t+o(t)}, \qquad x'(t)/x(t) \sim -1, \text{ as } t \to \infty.$$

Consider finally the case $\lambda_0 = 0$. Then (1) reads

$$x'' + f(t)x = 0. \tag{13}$$

Integration of (13) shows that the limit of $x'(t)$ exists as $t \to \infty$, and since x is of class $L^2(0, \infty)$, this limit must be 0. Therefore $x(t)$ is $o(1)$ as $t \to \infty$, which means that x'' and, in turn, x' is of class $L^2(0, \infty)$. This completes the proof of (IV).

6. The first assertion of (I) was proved in the lemma (IV). In order to prove (ii) and (iii) in (I), it can be assumed that $\lambda = 1$. Accordingly, if (1) has a solution $x(t) \not\equiv 0$ of class $L^2(0, \infty)$, then $x(t)$ satisfies (11). Hence (4) (which implies (7)), and (11) give

$$x^2(t) \leqq \int_t^\infty f^2(u)\,du \int_t^\infty x^2(v)\,dv. \tag{14}$$

Moreover, the solution $x(t)$ is assumed to be not identically zero, so that the positive function

$$r = r(t) = \int_t^\infty x^2(s)\,ds$$

is strictly decreasing. After introduction of r into (14) as a new dependent variable we obtain

$$0 \leqq -r' \leqq r \int_t^\infty f^2(s)\,ds,$$

that is,

$$0 \geqq (\log r)' \geqq -\int_t^\infty f^2(s)\,ds. \tag{15}$$

Upon integration of (15) there results

$$\log\,(r(t)/r(0)) \geqq -\int_0^t \int_\mu^\infty f^2(s)\,ds\,du. \tag{16}$$

Since $r(t) = o(1)$ as $t \to \infty$, (16) implies

$$\lim_{t\to\infty} \int_0^t sf^2(s)\,ds = \infty,$$

but this contradicts (4). Hence (ii) follows.

The truth of (iii) can now be concluded from an oscillation theorem of Hartman and Wintner which states that every λ is in S_α for some α, cf. [3]. Since no positive λ is in P_α for *any* α, it follows that every positive λ is in C_α for *some* α. On the other hand, every λ in C_α for some α, is in the essential

spectrum S'. Since S' is the derived set of S_a, every positive λ is in C_α for every α. The fact that C_α is a closed set completes the proof of (iii).

7. The assertion (iv) will be proved for (I), (II) and (III) together. It is known that any one of the conditions

$$f(t) = o(1) \text{ as } t \to \infty,$$
$$f(t) \text{ is of class } L^p(0, \infty),\ p \geqq 1$$

is sufficient in order that (1) be of Type I and that all solutions of (1) be non-oscillatory (that is, have only a finite number of zeros) when $\lambda < 0$ (see [1] and the references cited there.) It was also proved in [1] that if the solutions of (1) are non-oscillatory for $\lambda = \lambda_0$, then no $\lambda < \lambda_0$ is in the continuous spectrum and only a finite number of characteristic numbers are less than λ_0. This proves (iv), and in view of (ii) or (ii*), (v) and (vi) whenever the latter are asserted.

8. The assumption (5) of (II) implies

$$f(t) = o(1) \text{ as } t \to \infty. \tag{17}$$

It is known that if $f(t)$ satisfies (17) and if $\lambda > 0$, then any solution $x(t)$ of (1) has, on some half-line $t_0 < t < \infty$, the following properties (see for example [7]). There exists a sequence $t_0 < u_1 < u_2 < \cdots$ such that the distance $u_{n+1} - u_n$ tends, as $n \to \infty$, to a finite positive limit $(= \pi/\lambda^{\frac{1}{2}})$, and $x(t)$ is positive and concave, or negative and convex (from below) according as t is on the open intervals (u_1, u_2), (u_3, u_4), $(u_5, u_6), \cdots$ or (u_2, u_3), (u_4, u_5), $(u_6, u_7), \cdots$, the points $t = u_1, u_2, u_3, \cdots$ being both the zeros and the points of inflection of $x(t)$. Furthermore $x(t)$ is nowhere constant, and so there exists on each of the intervals (u_n, u_{n+1}) a unique point, say $t = v_n$, at which the derivative of $x(t)$ vanishes. The absolute value of the slope $x'(t)$ has on the interval (v_n, v_{n+1}) a unique maximum, say b_{n+1} and

$$b_{n+1} = x'(u_{n+1}). \tag{18}$$

Thus the local maxima of the absolute values of $x'(t)$ occur just at the zeros of $x(t)$. Moreover, let c_n denote the (absolute) area of the n-th half-wave of $x(t)$. Then it is known [7] that

$$c_{n+1}/c_n \to 1, \tag{19}$$

$$b_{n+1}/b_n \to 1, \tag{20}$$

and

(21) $$c_n/b_n \to 2/\lambda.$$

In order to prove the above facts Wintner made use of the inequality

(22) $$\log |(h(\beta)/h(\alpha))| \leqq \int_\alpha^\beta |f(t)|\, dt/\lambda^{\frac{1}{2}},$$

where $h = \frac{1}{2}(x'^2 + \lambda x^2)$, and α ($> t_0$) and $\beta > \alpha$ are arbitrary; [7], page 391.

It will be shown that Σc_n^2 diverges by virtue of the assumptions of (II), and this in turn will imply that $x(t)$ is not of class $L^2(0, \infty)$. In (22) put $\alpha = u_n$, $\beta = u_{n+1}$. Then since $x'(u_n) = b_n$, $x(u_n) = 0$,

(23) $$|\log (b^2_{n+1}/b^2_n)| \leqq \int_{u_n}^{u_{n+1}} |f(t)|\, dt/\lambda^{\frac{1}{2}}.$$

Let $\lambda > b^2$, and choose $\epsilon > 0$ so small that

$$0 < \lambda^{-\frac{1}{2}}(b+\epsilon)(\pi\lambda^{-\frac{1}{2}}+\epsilon)/(\pi\lambda^{-\frac{1}{2}}-\epsilon) \leqq 1.$$

Since $u_{n+1} - u_n = \pi\lambda^{-\frac{1}{2}} + o(1)$, there exists on the one hand a positive integer N such that

$$u_{n+1} - u_n < \pi\lambda^{-\frac{1}{2}} + \epsilon \text{ for } n > N,$$

and, on the other hand (by telescoping the reverse inequality),

$$u_n \geqq n(\pi\lambda^{-\frac{1}{2}} - \epsilon) \text{ for } n > N.$$

Choose N so large that also

$$t\,|f(t)| < b + \epsilon \qquad \text{for } t \geqq u_n,\ n > N.$$

Then

(24) $$\lambda^{-\frac{1}{2}} \int_{u_n}^{u_{n+1}} |f(t)|\, dt \leqq \lambda^{-\frac{1}{2}}(u_{n+1} - u_n) \max_{u_n \leqq t \leqq u_{n+1}} |f(t)|$$
$$\leqq \lambda^{-\frac{1}{2}}(\pi\lambda^{-\frac{1}{2}}+\epsilon)(b+\epsilon)/n(\pi\lambda^{-\frac{1}{2}}-\epsilon) \leqq 1/n$$

by virtue of the preceding inequalities. Combining (23) and (24) we obtain

(25) $$\log (b^2_{n+1}/b_n^2) \geqq -1/n \text{ for } n > N,$$

and (25) is sufficient that Σb_n^2 diverge. Hence, according to (21), Σc_n^2 also diverges. By Schwarz's inequality,

$$c_n^2 = \left(\int_{u_n}^{u_{n+1}} x(t)\, dt\right)^2 \leqq (u_{n+1} - u_n) \int_{u_n}^{u_{n+1}} x^2(t)\, dt.$$

The boundedness of $u_{n+1} - u_n$ and the divergence of Σc_n^2 now imply

$$\int^{\infty} x^2(t)\,dt = \infty.$$

This proves (ii*) of (II). Then (iii*) follows from (ii*) as did (iii) from (ii) in (I). Since (i), (iv), and (v) of (II) were proved in **7**, the proof of (II) is complete.

9. It is clear from the proofs of (I) and (II) that only assertion (ii) of (III) requires proof. If (a) holds and if $\lambda > 0$, then any solution $x(t)$ of (1) satisfies the well-known asymptotic relation

$$x(t) = a\cos\lambda^{\frac{1}{2}}t + \mathrm{b}\sin\lambda^{\frac{1}{2}}t + o(1) \quad \text{as } t \to \infty.$$

Hence $x(t)$ cannot be of class $L^2(0, \infty)$.

Much less than an asymptotic formula is needed for case (b) of (III). Let $\lambda > 0$ and let $u_1, u_2, \cdots, u_n, \cdots$ be the successive zeros of any solution $x(t)$ of (1). Then if $f(t)$ satisfies (b) there exists a constant $c > 0$ such that

$$\int_{u_n}^{u_{n+1}} x^2(t)\,dt = c + o(1) \qquad \text{as } n \to \infty,$$

cf. [6], pp. 258-260. Thus again $x(t)$ cannot be of class $L^2(0, \infty)$.

10. All the assertions of **4** can be proved by applying a method of construction due to Wintner [6]. Let μ and $\nu > 0$ be constants. Then

$$x(t) = \exp\left(\mu\int^{t} s^{-1}\cos^2\nu s\,ds\right)\cos\nu t, \qquad t > 0, \tag{26}$$

is a solution of

$$x'' + (\nu^2 + f(t))x = 0, \tag{27}$$

in which

$$f(t) = 2\mu\nu t^{-1}\sin 2\nu t + \mu t^{-2}\cos^2\nu t(1 - \mu\cos^2\nu t). \tag{28}$$

It is evident that the potential $f(t)$ defined in (28) does not satisfy (a) or (b), but does satisfy (a*) and (b*). Moreover, the boundedness of $tf(t)$ implies that (6) also holds. Further

$$\limsup t\,|\,f(t)\,| = 2\,|\,\mu\,|\,\nu. \tag{29}$$

Since $\int^{t} s^{-1}\cos^2\nu s\,ds \sim \frac{1}{2}\log t$, the solution $x(t)$ of (27) can, according to (26), be written

$$x(t) = t^{\mu/2+o(1)}\cos\ t.$$

If $\mu < -1$, then $x(t)$ is of class $L^2(0, \infty)$. The truth of (I^*), (II^*) and (III^*) is now easily inferred.

If we put $\nu = 0$ in (26), (27) and (28), we obtain the differential equation

$$x'' + \mu(1-\mu)t^{-2}x = 0$$

with the solution $x(t) = t^{\mu}$. This shows that the lower bound b^2 for λ can be in the point spectrum. It follows also that the limitation $\lambda_0 \neq 0$ in the last assertion of (IV) cannot be removed.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman, "On differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[2] ———, "Unrestricted solution fields of almost-separable differential equations," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 560-580.

[3] P. Hartman and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[4] E. C. Titchmarsh, *Eigenfunction Expansions*, Oxford, 1946.

[5] A. Wintner, "Small perturbations," *American Journal of Mathematics*, vol. 67 (1945), pp. 417-430.

[6] ———, "Asymptotic integrations of the adiabatic oscillator," *ibid.*, vol. 69 (1947), pp. 251-272.

[7] ———, "The adiabatic linear oscillator," *ibid.*, vol. 68 (1946), pp. 385-397.

[8] H. Weyl, "Uber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

THE SPECTRA OF PERIODIC POTENTIALS.*

By Sylvan Wallach.

1. Let $f(t)$ be a real-valued continuous function with the period 1, and let λ be a real parameter. Boundary value problems associated with the differential equation

$$(1) \qquad x'' + (\lambda + f(t))x = 0$$

have been extensively discussed in the mathematical and physical literature (see for example [3], [4]). The nature of the spectra for certain of these boundary value problems will be examined below.

It is well known that the characteristic equation of (1) has the form

$$(2) \qquad \rho^2 - 2A(\lambda)\rho + 1 = 0,$$

where $A(\lambda)$ depends only on λ. If complex values of the parameter λ are admitted in (1), then $A(\lambda)$ is an entire function which is real when λ is real.

Let $\rho(\lambda)$ be one of the roots of (2), so that the other root is $1/\rho(\lambda)$. Then there exist two real sequences λ_i, $i = 0, 1, 2, \cdots$, and λ_j^*, $j = 1, 2, \cdots$, such that (1) has a periodic solution $\phi_i(t)$ with period 1 when $\lambda = \lambda_i$, and a half-periodic solution $\phi_j^*(t)$ with period 2 when $\lambda = \lambda_j^*$. On the interval $0 \leqq t < 1$, $\phi_i(t)$ has i or $i+1$ zeros and $\phi_j^*(t)$ has $j-1$ or j zeros according as i and j are even or odd. Furthermore,

$$(3) \qquad \lambda_0 < \lambda_1^* \leqq \lambda_2^* < \lambda_1 \leqq \lambda_2 < \lambda_3^* \leqq \lambda_4^* < \cdots.$$

The corresponding values of ρ are $\rho(\lambda_i) = 1$, $\rho(\lambda_j^*) = -1$, and the λ-values occurring in (3) are the only ones for which $\rho(\lambda) = \pm 1$.

If $\rho(\lambda) \neq \pm 1$, that is, if $\lambda \neq \lambda_i$, $\lambda \neq \lambda_j^*$, then (1) possesses two linearly independent, complex-valued solutions of the forms

$$(4_1) \qquad x_1(t) = \rho^t \theta_1(t), \qquad\qquad (4_2) \qquad x_2(t) = \rho^{-t}\theta_2(t),$$

in which θ_1 and θ_2 have period 1, so that

$$(5_1) \qquad x_1(t+1) = \rho x_1(t), \qquad\qquad (5_2) \qquad x_2(t+1) = \rho^{-1} x_2(t).$$

It is further known that if λ is in one of the open intervals $(-\infty, \lambda_0)$, $(\lambda_{2i+1}, \lambda_{2i+2})$ or $(\lambda_{2j-1}^*, \lambda_{2j}^*)$, then $\rho(\lambda)$ is real, so that the solutions (4) are unbounded on $-\infty < t < \infty$. Such λ-values and the corresponding intervals

* Received April 22, 1948.

are said to be of unstable type. On the other hand, if λ is in one of the open intervals $(\lambda_{2i}, \lambda_{2i+1}^*)$ or $(\lambda_{2i+2}^*, \lambda_{2i+1})$, then $\rho(\lambda)$ is not real, in which case the solutions (4) are bounded—in fact, uniformly almost-periodic. Accordingly, these λ-values and intervals are said to be of stable type. For the facts stated above see, for example, [4], p. 15.

2. Let n be a positive integer. The differential equation (1), the boundary condition

$$x(0) = x(n) = 0, \tag{6}$$

and the normalization

$$\int_0^n x^2(t)\,dt = 1 \tag{7}$$

determine a point spectrum $S^n = \{\lambda_i^n\}$, $i = 0, 1, 2, \cdots$, and a sequence of characteristic functions ψ_i^n. The following facts concerning S^n will be proved.

(I) *The spectrum $S^n = \{\lambda_i^n\}$ of the boundary value problem determined by* (1), (6) *and* (7) *has the following properties:*

(i) *If $\rho(\lambda)$ is a complex $2n$-th root of unity, then λ is in S^n.*

(ii) *$\rho(\lambda_{mn-1}^n)$ is real for $m = 1, 2, \cdots$.*

(iii) *$\lambda_{mn-1}^n = \lambda_{m-1}^1$ for $n = 1, 2, \cdots$.*

(iv) *$\rho(\lambda_i^n)$ is a complex $2n$-th root of unity for $i \neq mn - 1$, $m = 1, 2, \cdots$. Consequently, there is just one characteristic number, λ_{mn-1}^n, in the closure of each interval of unstable type and there are $n - 1$ successive characteristic numbers, λ_i^n, $i = mn, \cdots, mn + n - 2$, in each open interval of stable type. Moreover, the characteristic functions $\psi_{mn-1}^n \equiv \psi_{m-1}^1$ have the form* (4_1) *or* (4_2). *All the other characteristic functions have a period, $2n$.*

Theorem (I) generalizes and completes a result of Wirtinger [7], who proved, for the case in which $f(t)$ is an *even* function, that λ is in S^n only if $\rho(\lambda)$ is a $2n$-th root of unity. Although Wirtinger does not explicitly assert the partial converse—that λ is in S^n if $\rho(\lambda)$ is a complex $2n$-th root of unity—its truth is evident from the arguments used. Wirtinger's proof is repeated, with somewhat more detail, by Hilb [2].

A corollary of (I) is:

(II) *A continuous determination of the argument of $\rho(\lambda)$ is a strictly monotone function of λ when $\rho(\lambda)$ is not real.*

Instead of (6) and (7), consider the boundary conditions

$$x(0)=0, \tag{8}$$

$$\int_0^\infty x^2(t)\,dt=1. \tag{9}$$

These boundary conditions, together with (1), define a boundary value problem and determine a spectrum in the sense of Hilbert's space. We shall prove:

(III) *The continuous spectrum S_c of the boundary value problem* (1), (8) *and* (9) *consists of the closure of the set of λ values of stable type; the point spectrum is contained in the set $\{\lambda_{m-1}{}^1\}$; and the cluster spectrum consists at most of the point $\lambda=\infty$. The characteristic functions are contained in the set $\{\psi_{m-1}{}^1\}$.*

The first assertion of (III) was proved by Wintner; cf. below.

(IV) *If $f(t)$ is an even function, then the point spectrum of the boundary value problem* (1), (8), (9) *is empty.*

The statement (IV) is particularly interesting in view of an oscillation theorem of Hartman and Wintner [1] which states that every λ not in the continuous spectrum or the cluster spectrum is in the point spectrum for some homogeneous boundary condition, $x(0)\sin\alpha+x'(0)\cos\alpha=0$. Thus, as α varies, the point spectrum sweeps over the intervals of unstable type. Nevertheless, (IV) states that the point spectrum is empty when $\alpha=\pi/2$.

3. It is not known in what sense, if any, the continuous spectrum S_c of the boundary value problem determined by (1), (8) and (9) can be defined to be the limit, as $T\to\infty$, of the spectra S^T of Sturm-Liouville boundary value problems associated with (1) on the bounded interval $0\leqq t\leqq T$. That some caution is required in formulating a definition, according to which S_c is the limit of S^T as $T\to\infty$, can be seen from the following simple considerations. Let the boundary condition (6) be replaced by

$$x(0)=x(T)=0, \qquad T>0, \tag{6 bis}$$

and let $S^T=\{\lambda_i{}^T\}$ be the corresponding point spectrum. It is then clear, from the continuous dependence of the points of the spectrum on T, that *every* $\lambda>\lambda_0$ is the limit of some sequence $\lambda_{i_k}{}^{T_k}$ in which i_k and T_k tend to ∞ with k. Moreover, if T is restricted to integral values, S_c is not the set of points, λ, for which there exist sequences $\lambda_{i_n}{}^n\to\lambda$ as $n\to\infty$, since $\lambda_{m-1}{}^1$ is in S^n for every n, but $\lambda_{m-1}{}^1$ is not, in general, in S^c. Nevertheless, the following statement can be proved.

(V) *For the boundary value problem* (1), (8) *and* (9), $S^n \to S_c$ *as* $n \to \infty$, *in the following sense: If* λ *is in* S_c, *then for every* $\epsilon > 0$ *and for every positive integer* k, *there exists an integer* $N(k, \epsilon)$ *such that the* ϵ-*neighborhood of* λ *contains more than* k *points of* S_n *whenever* $n > N(k, \epsilon)$; *on the other hand, if* λ *is not in* S_c, *there exists an* $\epsilon > 0$ *such that the* ϵ-*neighborhood of* λ *never contains more than one point of* S_n, $n = 1, 2, \cdots$.

According to (III), the continuous spectrum S_c consists of a sequence of closed intervals, viz. the closures of the intervals of stable type. The content of (V) can be expressed by saying that, as $n \to \infty$, the points of S^n become dense in the intervals of stable type, whereas the unstable intervals remain free of the point spectra except for isolated λ-values.

4. Theorems (III), (IV), and (V) complement certain results of Hilb [2] who obtained the analogue of the Fourier integral theorem for that case of (1) in which $f(t)$ is even. Hilb's method consists essentially of two steps—first representing the Green's function for the boundary condition (6) in terms of the characteristic functions and then passing to the limit as $n \to \infty$. He proves that the Green's functions tend to a bounded kernel which is expressed as an integral whose domain is the "continuous spectrum." The "continuous spectrum" was taken to be the sequence of intervals in which the points of the S^n become dense. That Hilb actually obtained the continuous spectrum in the sense of Hilbert's space is shown by (III) and (V).

5. Proof of (I). Let $\rho(\lambda) \neq \pm 1$. Then the general solution $x(t)$ of (1) is a linear combination,

$$x(t) = ax_1(t) + bx_2(t), \tag{10}$$

of the linearly independent solutions (4). In view of (5), the solution (10) satisfies (6) and (7) if and only if the system of homogeneous equations

$$\begin{aligned} x(0) &= ax_1(0) + bx_2(0) = 0 \\ x(n) &= a\rho^n x_1(0) + b\rho^{-n}x_2(0) = 0 \end{aligned} \tag{11}$$

possesses a non-trivial solution (a, b), that is, if and only if the determinant,

$$\Delta = x_1(0)x_2(0)(\rho^{-n} - \rho^n), \tag{12}$$

of (11) vanishes. According to (12), $\Delta = 0$ whenever $\rho(\lambda)$ is a $2n$-th root of unity. This proves (i).

Suppose, on the other hand, that $\lambda = \lambda_i^n$ is a characteristic number with $\rho(\lambda)$ *not* a $2n$-th root of unity. Then it is evident from (12) that $\Delta = 0$ only if $x_1(0)$ or $x_2(0)$ vanishes, say $x_1(0) = 0$. From the form (4_1) of $x_1(t)$

it is clear that the zeros of $x_1(t)$ are periodic, with period 1. Hence $x_1(n)=0$. It follows that the characteristic function $\psi_i{}^n$ is some constant multiple (perhaps complex) of $x_1(t)$. Separation of $x_1(t)$ into real and imaginary parts shows that this is possible only if ρ is real.

Moreover, $\psi_i{}^n$ vanishes at $t=0,1,2,\cdots$, so that $\psi_i{}^n=\psi_{j_m}{}^m$, $m=1,2,\cdots$. In other words, every characteristic number λ for which $\rho(\lambda)\neq\pm1$ is real is a characteristic number for every n. That this is true also for $\rho(\lambda)=\pm1$ can be seen as follows. Let $\delta=0$ or 1 according as ρ is $+1$ or -1. Then, when $\rho=\pm1$, there exist linearly independent solutions $e^{i\pi\delta t}\chi_1(t)$, and $e^{i\pi\delta t}(\chi_2(t)+ct\chi_1(t))$, in which χ_1 and χ_2 have period 1, and c is a constant. If

$$\psi_i{}^n(t)=e^{i\pi\delta t}\{a\chi_1+b(\chi_2+ct\chi_1)\}$$

is a characteristic function, vanishing at $t=0$, $t=n$, it is easily seen that $b=0$. Hence $\psi_i{}^n$ vanishes at $t=0,1,2,\cdots$.

So far we have proved that:

(*) *When $\rho(\lambda)$ is not real, λ is a characteristic number if and only if $\rho(\lambda)^{2n}=1$; when $\rho(\lambda)$ is real, λ is a characteristic number for every n or for no n.*

6. It will now be proved that there is just one characteristic number in the closure of each interval of unstable type, $(-\infty,\lambda_0)$ excluded. According to (*), this assertion is independent of n; hence it may be assumed that $n=1$. But then there are no characteristic numbers for which $\rho(\lambda)$ is complex, and the assertion follows from the remark preceding (3) by virtue of Sturm's comparison theorem.

Consider now the sequence of characteristic numbers

$$\lambda_0{}^1,\lambda_1{}^1,\lambda_2{}^1,\cdots,\lambda_{m-1}{}^1,\cdots,$$

and the corresponding sequence of characteristic functions

$$\psi_0{}^1,\psi_1{}^1,\psi_2{}^1,\cdots,\psi_{m-1}{}^1,\cdots.$$

Since $\psi_{m-1}{}^1$ has periodic zeros, vanishes at $t=0,1,2,\cdots$, and has m zeros on $[0,1]$, $\psi_{m-1}{}^1$ has $mn+1$ zeros on $[0,n]$. Consequently,

$$\lambda_{m-1}{}^1=\lambda_{mn-1}{}^n;\quad \psi_{m-1}{}^1=\psi_{mn-1}{}^n,\qquad n=1,2,\cdots, \tag{13}$$

and the characteristic numbers in (13) are *all* the characteristic numbers for which $\rho(\lambda)$ is real. This proves (ii) and (iii).

For fixed n, the characteristic numbers occurring in (13) are

$$\lambda_{n-1}{}^{n}, \lambda_{2n-1}{}^{n}; \lambda_{3n-1}{}^{n}, \cdots.$$

Hence there are just $n-1$ characteristic numbers in each open interval of stable type. This proves (iv) and completes the proof of (I).

7. Proof of (II). It is known, cf. e. g. Strutt, *loc. cit.*, that $\rho(\lambda)$ is positive in the intervals $\lambda_{2i+1}, \lambda_{2i+2}$ and negative in the intervals $\lambda_{2j-1}{}^{*}, \lambda_{2j}{}^{*}$. According to (2), the continuous function $\rho(\lambda)$ cannot change sign without crossing one-half of the complex unit circle. Thus, as λ traverses an interval of stable type, $\rho(\lambda)$ is a complex $2n$-th root of unity at least $n-1$ times. The truth of (II) can now be inferred from the facts that there are exactly $n-1$ characteristic numbers in an open interval of stable type, the $2n$-th roots of unity are dense in the unit circle, and $\rho(\lambda)$ is an analytic function.

8. Proofs of (III) and (IV). The differential equation (1) has a solution $x(t) \not\equiv 0$ of class $L^2(0, \infty)$ if and only if λ is of unstable type. Let $x(t) \not\equiv 0$ be of class $L^2(0, \infty)$. Then if (8) is also satisfied, $x(t) = 0$ for $t = 0, 1, 2, \cdots$, and so $x(t)$ also satisfies (6) and, except for a constant factor, (7). Therefore, $\lambda = \lambda_{m-1}{}^{1}$ for some m. This proves all of (III) except the assertion concerning the continuous spectrum. This assertion, however, is a particular case of a theorem of Wintner on the location of continuous spectra, [5] p. 23, see also [6].

In order to prove (IV), one need only observe that according to Wirtinger, *loc. cit.*, $f(t) = f(-t)$ implies that $\rho(\lambda)^{2n} = 1$ for *every* characteristic number λ of the boundary value problem (1), (6). But then every characteristic function is either uniformly almost periodic or of the form $\chi_2(t) + ct\chi_1(t)$. Hence no solution can be of class $L^2(0, \infty)$.

9. Proof of (V). The proof of (V) is implicit in the arguments used above. If λ is in S_c, then every neighborhood of λ contains infinitely many λ values for which $\rho(\lambda)$ is a complex $2n$-th root of unity. On the other hand, in the interior of each interval of unstable type there is at most one λ value, namely $\lambda_{m-1}{}^{1}$, which can be in S^n.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[2] E. Hilb, "Ueber Integraldarstellung willkürlicher Funktionen," *Mathematische Annalen*, vol. 66 (1909), pp. 1-66.

[3] F. Seitz, *The Modern Theory of Solids*, New York, 1940.

[4] M. J. O. Strutt, *Lamésche, Mathieusche und verwandte Funktionen in Physik und Technik*, Berlin, 1932.

[5] A. Wintner, "On the location of continuous spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 22-30.

[6] ———, "Stability and spectrum in the wave mechanics of lattices," *Physical Review*, vol. 72 (1947), pp. 81-82.

[7] W. Wirtinger, "Beiträge zu Riemann's Integrationsmethode für hyperbolische Differentialgleichungen, und deren Anwendung auf Schwingungsprobleme," *Mathematische Annalen*, vol. 48 (1897), pp. 365-389.

THE LEAST CLUSTER POINT OF THE SPECTRUM OF BOUNDARY VALUE PROBLEMS.*

By PHILIP HARTMAN and CALVIN R. PUTNAM.

1. In the differential equation

$$y'' + (\lambda - q)y = 0, \tag{1}$$

let $q = q(x)$ be a real, continuous function on $0 \leqq x < \infty$ and let λ denote a real parameter. Only real-valued solutions of (1) will be considered. For a fixed λ, the differential equation (1) is said to be oscillatory or non-oscillatory according as every non-trivial solution $y = y(x)$ has or does not have an infinity of zeros on $0 \leqq x < \infty$.

If, for some λ (and hence for all λ), (1) possesses a solution $y = y(x)$ for which

$$\int_0^\infty y^2(x)\,dx = \infty \tag{2}$$

is satisfied, the differential equation (1) is in the *Grenzpunktfall* according to Weyl's terminology; [7], p. 238. In this case, (1) and a homogeneous boundary condition

$$\alpha y(0) + \beta y'(0) = 0, \qquad (\alpha^2 + \beta^2 \neq 0) \tag{3}$$

determine an eigenvalue problem. The set S' of cluster points λ of the spectrum of the boundary value problem (1), (3) is independent of the choice of the boundary condition (3) and is determined by (1) alone; [7], p. 251. The object of this paper is to characterize the least point (possibly $\pm\infty$) of the closed set S'.

For example, it is known that if $q = q(x)$ satisfies the condition

$$\mu = \liminf_{x\to\infty} q(x) > -\infty, \tag{4}$$

then the differential equation (1) is in the Grenzpunktfall; [7], p. 238; furthermore, if (1) is oscillatory for $\lambda = \mu$, then $\lambda = \mu$ is the least point of S'; [7], p. 252.

Also, it has recently been shown [6], § 10 that if

$$\mu = \lim_{x\to\infty} q(x) \tag{5}$$

* Received April 27, 1948.

exists, then $\lambda = \mu$ is the least point of S' (whether or not (1) is oscillatory when $\lambda = \mu$). The proof of this assertion depends on an adaptation of the variational methods associated with non-singular Sturm-Liouville boundary value problems. This method will be combined with recent results [4] on differential equations (1) which are non-oscillatory for some λ to prove

(*) *Let $q = q(x)$, where $0 \leqq x < \infty$, be continuous. Define μ as follows:*

(i) *if* (1) *is non-oscillatory for all λ, let $\mu = \infty$;*

(ii) *if* (1) *is non-oscillatory for some but not all λ, let μ be the unique number with the property that* (1) *is oscillatory when $\lambda > \mu$ and non-oscillatory when $\lambda < \mu$;*

(iii) *if* (1) *is oscillatory for all λ, let $\mu = -\infty$.*

In the cases (i)-(ii), *the differential equation* (1) *is in the Grenzpunktfall and μ is the least point of S'; in the case* (iii), *μ is the least point of S' whenever* (1) *is in the Grenzpunktfall.*

The cases (i) and (iii) of (*) are known; cf. [4], Corollary, § 6, and [11], pp. 313-314, respectively. That (1) is in the Grenzpunktfall in the case (ii) is also known; [4], Theorem, § 1.

If (1) is replaced by

(1 bis) $$(py')' + (\lambda - q)y = 0,$$

in (*), where $p = p(x) > 0$ and $q = q(x)$ are continuous and such that (1 bis) is in the Grenzpunktfall, then the proof of (*) will show that the corresponding number μ is the least cluster point of the spectrum of the boundary value problem determined by (1 bis) and (3). (It can be mentioned that if (1 bis) is non-oscillatory for some λ, then the condition

$$\int_0^\infty p^{-\frac{1}{2}}(x)\,dx = \infty$$

is sufficient for (1 bis) to be in the Grenzpunktfall; [4], end of § 2.)

Clearly, the theorem (*) contains the assertions made above concerning (4) and (5). Also, (*) and Sturm's comparison theorem imply

COROLLARY 1. *If $q = q(x)$, where $0 \leqq x < \infty$, is a continuous, bounded function, then* (1) *is in the Grenzpunktfall and the least point μ of S' satisfies*

$$\liminf_{x\to\infty} q(x) \leqq \mu \leqq \limsup_{x\to\infty} q(x).$$

Another consequence of (*) and Sturm's comparison theorem is

COROLLARY 2. *Let $q_1(x)$ and $q_2(x)$ be two continuous functions on the half-line $0 \leqq x < \infty$ such that*

$$q_1(x) - q_2(x) \to 0 \text{ as } x \to \infty. \tag{6}$$

Then either both of the differential equations

$$y'' + (\lambda - q_k)y = 0, \qquad (k = 1, 2), \tag{7_k}$$

are or neither of them is in the Grenzpunktfall; in the first case, the spectra of the boundary value problems determined by (3) *and* (7_k), *for $k = 1$ or 2, have the same least cluster point (possibly $\pm \infty$).*

The first assertion of this corollary, that concerning the Grenzpunktfall, is known and holds even if (6) is replaced by

$$| q_1(x) - q_2(x) | < \text{const.} \qquad (0 \leqq x < \infty);$$

cf. [1], p. 513 or [9], p. 266.

It remains undecided whether or not the "same least cluster point" in the last part of Corollary 2 can be replaced by the "same set of cluster points."

Finally, it will be shown that (*) implies

COROLLARY 3. *If the continuous function $q = q(x)$, where $0 \leqq x < \infty$, satisfies*

$$\int_0^\infty | q(x) - \mu |^r \, dx < \infty \text{ for some } r \geqq 1 \tag{8}$$

and some number μ, then (1) *is in the Grenzpunktfall and μ is the least point of S'.*

The case $p = 1$ of Corollary 3 follows from a general oscillation theorem [5] and known asymptotic formulae for the solutions of (1). (The required asymptotic formulae can be obtained from [2] by a suitable change of variables; cf. also [8] and [10].) This particular case of Corollary 3 can also be obtained from [7], p. 258, if it is noted that the assumption (5) is actually not needed in the proof of the theorem [7], p. 258.

2. Proof of (*). Since the cases (i) and (iii) of (*) are known, the case (ii) will first be proved. In this case, it only remains to show that μ is the least point of S'.

The theorem [4], § 1, implies that if (1) is oscillatory when $\lambda = \mu$, then $\lambda = \mu$ is the least point of S'. On the other hand, if (1) is non-oscillatory when $\lambda = \mu$, then the spectrum of the boundary value problem determined by

(1) and (3) contains at most a finite number, say N, of points which do not exceed μ. Let

$$\lambda_1 < \lambda_2 < \cdots < \lambda_N, \qquad (0 \leqq N < \infty;\ \lambda_N \leqq \mu),$$

denote these points.

Suppose, if possible, that μ is not a cluster point of the spectrum. Then there exists a number $\epsilon > 0$ such that the λ-interval $\mu < \lambda \leqq \mu + 2\epsilon$ does not contain a point of the spectrum.

Let $\phi_n = \phi_n(x)$, where $n = 1, 2, \cdots, N$, denote the normalized eigenfunctions belonging to $\lambda_1, \lambda_2, \cdots, \lambda_N$ respectively; so that $y = \phi_n(x)$ is the solution of (1), with $\lambda = \lambda_n$, satisfying (3) and

$$\int_0^\infty y^2(x)\,dx = 1. \tag{9}$$

Let $\phi = \phi(x)$ denote an arbitrary non-trivial solution of (1), where $\lambda = \mu + \epsilon$. In view of the definition of μ the function $\phi(x)$ possesses a sequence of zeros $x_0 < x_1 < x_2 < \cdots$. In terms of a set of numbers $c_1, c_2, \cdots, c_{N+1}$, define a continuous function $\Phi = \Phi(x)$ for $0 \leqq x < \infty$ by the formulae

$$\begin{aligned} &\Phi(x) = c_j\phi(x) \text{ for } x_{j-1} \leqq x < x_j \text{ and } j = 1, \cdots, N+1; \\ &\Phi(x) = 0 \text{ for } 0 \leqq x < x_0 \text{ and } x_{N+1} \leqq x < \infty. \end{aligned} \tag{10}$$

The set of orthogonality relations

$$\int_0^\infty \Phi(x)\phi_n(x)\,dx = 0 \text{ for } n = 1, \cdots, N \tag{11}$$

is either vacuous or is a set of N linear homogeneous equations for the $N+1$ constants $c_1, \cdots, c_{N+1}$. It can therefore be assumed that these constants have been chosen so as to satisfy (11), and the normalization

$$\int_0^\infty \Phi^2(x)\,dx = 1. \tag{12}$$

The definition (10) shows that

$$\int_0^\infty (\Phi'^2 + q\Phi^2)\,dx = \sum_{j=1}^{N+1} c_j^2 \int_{x_{j-1}}^{x_j} (\phi'^2 + q\phi^2)\,dx.$$

An integration by parts gives

$$\int_{x_{j-1}}^{x_j} (\phi'^2 + q\phi^2)\,dx = -\int_{x_{j-1}}^{x_j} \phi(\phi'' - q\phi)\,dx,$$

for $\phi(x_j)=\phi(x_{j-1})=0$. Since $\phi(x)$ is a solution of (1) with $\lambda=\mu+\epsilon$, the last integral equals

$$(\mu+\epsilon)\int_{x_{j-1}}^{x_j}\phi^2dx.$$

Hence,

$$\int_0^\infty(\Phi'^2+q\Phi^2)\,dx=(\mu+\epsilon)\sum_{j=1}^{N+1}c_j{}^2\int_{x_{j-1}}^{x_j}\phi^2dx=(\mu+\epsilon)\int_0^\infty\Phi^2dx,$$

so that, by (12),

$$\int_0^\infty(\Phi'^2+q\Phi^2)\,dx=\mu+\epsilon.$$

Let $\rho(\lambda)$, where $-\infty<\lambda<\infty$, be a monotone function which determines the spectral resolution for the boundary value problem (1) and (3); so that the spectrum consists of those points λ, in every neighborhood of which $\rho(\lambda)$ is not constant. Cf. Weyl [7], p. 239; in contrast to the notation of Weyl, it is assumed here that the monotone function has appropriate discontinuities at the eigenvalues so that the standard relations involving "Fourier transforms" can be expressed in terms of a Hellinger integral only, instead of a series and a Hellinger integral; see, e. g., [6], § 1-§ 2.

Let $\Gamma(\lambda)$, $-\infty<\lambda<\infty$, be the "Fourier transform" of $\Phi(x)$, so that

$$\int_{-\infty}^\infty(d\Gamma)^2/d\rho=\int_0^\infty\Phi^2(x)\,dx=1;$$

as above, the notation differs from that of Weyl [7], p. 251, in that the contributions of the eigenfunctions as well as the eigendifferentials are contained in $\Gamma(\lambda)$. Then it follows from the Lemma in § 2, [6], and the Corollary and Remark in § 3, [6], that

$$\int_0^\infty(\Phi'^2+q\Phi^2)\,dx\geqq\int_{-\infty}^\infty\lambda(d\Gamma)^2/d\rho.$$

Since the only contribution of the half-line $-\infty<\lambda\leqq\mu+2\epsilon$ to the spectrum is the set of eigenvalues $\lambda_1,\cdots,\lambda_N$ and since $\Phi(x)$ satisfies the orthogonality relations (11), it follows that $d\Gamma(\lambda)=0$ for $\lambda<\mu+2\epsilon$ and so

$$\int_{-\infty}^\infty\lambda(d\Gamma)^2/d\rho\geqq(\mu+2\epsilon)\int_{-\infty}^\infty(d\Gamma)^2/d\rho.$$

The last four formula lines lead to the contradiction $\mu+\epsilon\geqq\mu+2\epsilon$. Therefore the assumption that $\lambda=\mu$ is not a cluster point of the spectrum is untenable. This completes the proof of the case (ii) of (*).

A proof similar to that above can also be used for the case (iii). Let (1) be in the Grenzpunktfall and oscillatory for every λ. Suppose, if possible, that $\lambda = -\infty$ is not the least cluster point of the spectrum of the boundary value problem determined by (1) and (3). Then there exists a pair of numbers μ^* and $\epsilon > 0$ such that the half-line $-\infty < \lambda \leqq \mu^* + 2\epsilon$ contains no point of the spectrum. Since (1) is oscillatory for $\lambda = \mu^* + \epsilon$, the proof can be completed as above (for the situation when $N = 0$). This completes the proof of the case (iii) and of (*).

3. Proof of Corollary 3. In order to prove this corollary, it is sufficient to show that if $q(x)$ and μ satisfy (8), then the case (ii) of (*) applies. It is known ([3], § 18) that (8) implies that (1) is non-oscillatory when $\lambda < \mu$. It only remains, therefore, to verify the fact that (1) is oscillatory when $\lambda > \mu$.

Suppose the contrary, so that there exists a $\lambda > \mu$ corresponding to which every solution $y = y(x)$ of (1) possesses at most a finite number of zeros. Let X be chosen so large that $y(x) \neq 0$ for every $x \geqq X$. If (1) is divided by $y = y(x)$, where $x \geqq X$, and the identity

$$y''/y = (y'/y)' + (y'/y)^2$$

is applied, it is seen that

$$(y'/y)' + (y'/y)^2 + (\lambda - \mu) + (\mu - q) = 0.$$

A quadrature of this relation leads to

$$y'/y = \text{const.} - \int_X^\infty (y'/y)^2 dx - (\lambda - \mu)x - \int_0^\infty (\mu - q)\, dx.$$

In view of (8) and the Hölder inequality

$$\left| \int_0^\infty (\mu - q)\, dx \right| \leqq x^{1-1/r} \left(\int_0^\infty |\mu - q|^r \, dx \right)^{1/r},$$

it follows from $\lambda - \mu > 0$ and $r \geqq 1$ that

$$y'/y \to -\infty \text{ as } x \to \infty.$$

In particular, y and y' are of opposite sign for large x, so that y tends monotonously to a finite limit. Consequently, y' is improperly integrable and therefore absolutely improperly integrable (since it cannot change sign for large x-values)

Thus, if $y = y(x)$ is an arbitrary solution of (1), the function $y = y(x)$ is bounded and its derivative $y'(x)$ is absolutely integrable. This, however,

contradicts the fact that the Wronskian of two linearly independent solutions of (1) is a non-vanishing constant and establishes Corollary 3.

This proof shows that Corollary 3 remains true when (8) is replaced by the Tauberian condition

$$\int_0^x (\mu - q(t))\,dt/x \to 0 \text{ as } x \to \infty.$$

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] R. Bellman, "A stability criterion of solutions of linear differential equations," *Duke Mathematical Journal*, vol. 11 (1944), pp. 513-516.

[2] M. Bôcher, "On regular singular points of linear differential equations of the second order whose coefficients are not necessarily analytic," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 40-52.

[3] P. Hartman, "Unrestricted solution fields of almost-separable differential equations," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 560-580.

[4] ———, "On differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[5] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[6] C. R. Putnam, "An application of spectral theory to a singular calculus of variations problem," *American Journal of Mathematics*, vol. 70 (1948), pp. 780-803.

[7] H. Weyl, "Ueber gewöhnliches Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlichen Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

[8] A. Wintner, "Small perturbations," *American Journal of Mathematics*, vol. 67 (1945), pp. 417-430.

[9] ———, "Asymptotic integrations of the adiabatic oscillator," *American Journal of Mathematics*, vol. 69 (1947), pp. 251-272.

[10] ———, "Asymptotic integrations of the adiabatic oscillator in its hyperbolic range," *Duke Mathematical Journal*, vol. 15 (1948), pp. 55-67.

[11] ——— and P. Hartman, "On the orientation of unilateral spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 309-316.

TWO-DIMENSIONAL TRANSONIC FLOW PATTERNS.* [1]

By Stefan Bergman.

1. Introduction. The successful development of the mathematical theory of two-dimensional steady flows of an ideal incompressible fluid is largely due to the fact that the complex potentials of such flows are analytic functions of a complex variable. Function-theoretical methods may therefore be used in order to obtain and to investigate flow patterns possessing at given points the assigned singular character and satisfying given boundary conditions.[2]

In investigating the flow of an incompressible fluid by theoretical methods, two alternative treatments have been used. The behavior of the solutions has been studied either in the so-called physical, i. e., the plane in which the motion actually occurs, or in the hodograph plane, i. e., the plane whose cartesian coordinates are the velocity components. In the incompressible case, both the potential and the streamfunction are harmonic functions, irrespective of whether the motion is considered in the physical—or the hodograph plane.

In the compressible case, both the potential and the streamfunction considered in the physical plane satisfy complicated non-linear equations. By considering the motion in the hodograph plane, and making a few appropriate transformations, it is possible to linearize these equations. (See [11] and **2** of the present paper.[3])

The treatment of the problem in the physical plane has, of course, the fundamental advantage of making the boundary conditions enter in an obvious way. The gain in simplicity due to the linearization is, however, so considerable as to make the hodograph method vastly superior for various purposes.

In the present paper the theory of compressible fluid flow is developed

* Received August 2, 1947.

[1] Research paper done under Navy Contract NOrd 8555-Task F, at Harvard University. The ideas expressed in this paper represent the personal views of the author, and are not necessarily those of the Bureau of Ordnance.

[2] Poles, logarithmic singularities, etc., of the complex potential represent doublets, sinks, sources, vortices, etc., of the flow.

[3] The numbers in brackets refer to the bibliography at the end of the paper. Acquaintance with the contents of these publications is not assumed in the present paper.

by studying the linearized equation and applying to it the operator method developed in [**1-10**]. Generalizing the procedure: "taking the real part," this method leads to the determination of certain linear operators which transform functions of one variable into stream (or potential) functions of compressible fluid flows preserving many fundamental properties of functions to which the operator is applied. In the present paper, two operators of this kind are considered: the so-called integral operator of the first kind (**3**) and that of the second kind (**4-6**). For some fluid dynamical applications, and in particular for the study of the behavior of the flow near the sonic line, the second operator seems to be more appropriate. In this case, however, the relations between the properties of functions to which the operator is applied and those of generated functions is more hidden. This is why it is useful at first to investigate the special case which is obtained by assuming a simplified equation of state. This simplifying assumption results in the equation for the stream function taking the form

$$(1.1) \qquad -C\mathrm{H}(\partial^2\psi/\partial\theta) + (\partial^2\psi/\partial\mathrm{H}^2) = 0, \qquad C > 0$$

where H is a function of the Mach number, M, which is negative for $M < 1$ and positive for $M > 1$. (See (2.1) and (2.3)).

In an important investigation, Tricomi [**16**] studied the boundary value problem of equation (1.1) and showed that if we consider a finite domain D, bounded in the supersonic region by two characteristics, say BA and CA, and by a curve BmC in the subsonic region, and if the boundary values are prescribed in BmC and on one of the characteristics, say BA, then the boundary value problem has a unique solution. Frankl [**13**] considered questions allied with Tricomi's investigations in the case of the exact compressibility equation.

The questions arising in our approach are, however, of a somewhat different nature than those considered by the above-mentioned authors.

In the first place, we seek to find conditions for a function, say f, of one variable in order that the generated function $\mathrm{P}(f)$ will be defined in a prescribed domain B, which in general lies partially in the subsonic and partially in the supersonic region.

Secondly, and this is the most essential difference, we are considering solutions of the compressibility equation which possess singularities (e. g., branchpoints) in the hodograph plane. In the applications of the theory, the consideration of this type of solution cannot be dispensed with, since, notwithstanding the singularities in the hodograph plane, the behavior of the solutions in the physical plane can be perfectly regular. Furthermore, certain

singularities in the physical plane have a hydrodynamical meaning and must be considered in investigating flow patterns.

The methods employed in the case of equation (1. 1) are to some extent capable of generalization to the case in which the coefficient $(-C\mathrm{H})$ is replaced by an arbitrary function $l(\mathrm{H})$. (See **5**). This includes, in particular, the exact, i. e., non-simplified compressibility equation.

In **3**, we introduce the so-called integral operator of the first kind, P_1. This operator yields a streamfunction of a subsonic, compressible fluid flow in terms of an arbitrary function of one complex variable. The representation holds for[4] $\mathrm{E}[M < 1, -\infty < \theta < \infty]$. Here M is the Mach number and θ the angle which the velocity vector forms with the positive x-axis.

An analogous representation for the streamfunctions of supersonic flows in terms of two differentiable functions of one real variable holds for $\mathrm{E}[M > 1, -\infty < \theta < \infty]$. Using the integral operator of the second kind, we obtain (**4**, **5**) four analogous representations in terms of arbitrary functions of one variable. These four representations are valid in four adjacent domains of the M, θ-plane, namely

$$D_1 = \mathrm{E}[M < 1, \theta > 3^{\frac{1}{2}} \,|\, \lambda(M) \,|] + \mathrm{E}[M > 1, \theta > \Lambda(M)],$$
$$D_2 = \mathrm{E}[M < 1, |\, \theta \,| < 3^{\frac{1}{2}} \,|\, \lambda(M) \,|],$$
$$D_3 = \mathrm{E}[M < 1, \theta < -3^{\frac{1}{2}} \,|\, \lambda(M) \,|] + \mathrm{E}[M > 1, \theta < -3\Lambda(M)],$$
$$D_4 = \mathrm{E}[M > 1, -3\Lambda(M) < \theta < \Lambda(M)],$$

respectively. Here $\theta = \pm 3^{\frac{1}{2}}\lambda(M)$ and $\theta = (-1 \pm 2)\Lambda(M)$ are certain curves which pass through the point $M = 1$, $\theta = 0$ and which lie in the subsonic and supersonic region respectively (see fig. 1).

In the simplified case these four representations can be combined into one, yielding a representation which holds in the whole M, θ-plane. This result is based on certain theorems of the Fuchs theory of ordinary differential equations with singular coefficients. The question of combining the analogues of the above four solutions in the exact case, and generally, the study of the solutions leads to the investigation of *partial* differential equations with singular coefficients, which, when solutions are continued to complex values of the arguments, can be attacked by methods representing a generalization

[4] The functions $\psi = \mathrm{P}_1(f)$ may be multi-valued functions which may possess singularities. The statement that $\mathrm{P}_1(f)$ is defined in $\mathrm{E}[M < 1]$ means that the projections of the domain in which $\mathrm{P}_1(f)$ is defined on the schlicht M, θ-plane, lies in $\mathrm{E}[M < 1]$.

$\mathrm{E}[\ \ ]$ denotes the set of points whose coordinates satisfy conditions indicated in the brackets.

of the Fuchs theory for ordinary differential equations. These questions, and in particular the problem of combining the four above representations into one, will be treated in a subsequent paper.

In **6**, we determine the "associate" function for $P_2(f)$ in terms of the values of the stream function $\psi = \mathrm{Im}[P_2(f)]$, and its derivative with respect to M on the line $M=1$ (sonic line).

The author would like to take this opportunity to thank Bernard Epstein and A. Zeichner for helpful advice and aid in connection with the present

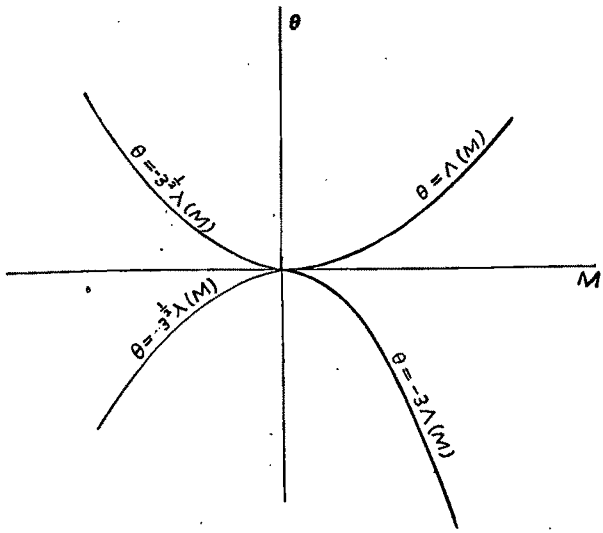


Fig. 1.

paper. He would also like to thank Z. Nehari and M. Schiffer for a number of helpful suggestions.

2. Equation for the streamfunction of a compressible fluid flow. Exact and simplified equations. Assuming that the thermodynamical equation of state of the fluid has the form $p = \sigma\rho^k$, where σ and k are constants, and ρ and p the density and pressure respectively, and introducing as new variables

$$\text{(2.1)} \qquad \mathrm{H} = \int_{q_1}^{q} \rho(dq/q) = \int_{q_1}^{q} q^{-1}[1 - \tfrac{1}{2}(k-1)q^2]^{1/k-1}dq,$$

and θ, where q is the speed and θ the angle which the velocity vector makes with some fixed direction (say the positive x-axis,), we obtain the following linear equation for the streamfunction:

$$(2.2)\quad S(\psi) \equiv l(\mathrm{H})(\partial^2\psi/\partial\theta^2) + (\partial^2\psi/\partial \mathrm{H}^2) = 0, \qquad l(\mathrm{H}) = (1 - M^2)/\rho^2,$$

where

$$(2.3)\qquad M = q/[1 - \tfrac{1}{2}(k-1)q^2]^{\frac{1}{2}}$$

is the Mach number. The denominator in (2.3), being the local velocity of sound, M will be smaller or larger than 1 according as the flow is subsonic or supersonic respectively. The differential equation (2.2) will therefore be of ellipitic or hyperbolic type, corresponding to the subsonic or supersonic character of the flow.

A formal computation [5] shows that the Taylor development of $l(\mathrm{H})$ in the neighborhood of $\mathrm{H} = 0$ is

$$(2.4)\quad l(\mathrm{H}) = [2/(k-1)]^{(2-k)/(k-1)}[(-2\mathrm{H}) - ((2k+5)/(2k+2))((k+1)/2)^{2k/(k-1)}(-2\mathrm{H})^2 + \frac{k^4 + (43/6)k^3 + 16k^2 + (31/2)k + 31/6}{(k+1)^4}((k+1)/2)^{k/(k-1)}(-2\mathrm{H})^3 + \cdots]$$

In considering a flow (or a portion of a flow) which is purely subsonic, it has some advantages to replace H by the variable λ defined by

$$(2.5)\qquad -\lambda = \int_{\tau=0}^{-\mathrm{H}} [l(-\tau)]^{\frac{1}{2}} d\tau.$$

λ can be expressed in a closed form as a function of M; a formal computation yields

$$(2.6)\qquad \lambda = \tfrac{1}{2}\log\left[\frac{1-(1-M^2)^{\frac{1}{2}}}{1+(1-M^2)^{\frac{1}{2}}}\left(\frac{1+h(1-M^2)^{\frac{1}{2}}}{1-h(1-M^2)^{\frac{1}{2}}}\right)^{1/h}\right],$$

$$h = [(k-1)/(k+1)]^{\frac{1}{2}}.$$

Now

$$-\psi_\lambda\lambda_{\mathrm{H}} = -\psi_\lambda[il(\mathrm{H})]^{\frac{1}{2}},$$

$$(2.7)\quad \psi_{\mathrm{HH}} = \psi_{\lambda\lambda}l + \psi_\lambda(l^{\frac{1}{2}})_{\mathrm{H}} = \psi_{\lambda\lambda}l - \tfrac{1}{2}\psi_\lambda l^{-\frac{1}{2}}l_{\mathrm{H}} = l[\psi_{\lambda\lambda} + \tfrac{1}{2}l^{-3/2}l_{\mathrm{H}}]$$

so that (2.2) becomes

$$(2.8a)\qquad \psi_{\lambda\lambda} + \psi_{\theta\theta} + 4N\psi_\lambda = 0,$$

[5] A detailed account of formal derivations of some expressions used in the present paper can be found in the Appendix to Technical Report 10, of the series "Operator methods in the theory of compressible fluids," Harvard University, 1948.

or, in complex notation,

(2.8b) $$4\psi_{Z\bar{Z}} + 4N(\psi_Z + \psi_{\bar{Z}}) = 0,$$

where

(2.9) $$N = \tfrac{1}{8}l^{-3/2}l_{\mathrm{H}} = -[(k+1)/8]\frac{M^4}{(1-M^2)^{3/2}}, \quad Z = \lambda + i\theta, \bar{Z} = \lambda - i\theta.$$

See [**6**, (46)]. It should be noted that the interval $-\infty < \lambda < 0$ corresponds to the interval $-\infty < \mathrm{H} < 0$.

In the supersonic case (i.e., for $M > 1$) the right-hand side of (2.6) becomes purely imaginary. If we introduce the new variable Λ defined by

(2.10) $$\Lambda = i\lambda,$$

it is easily confirmed that

(2.11) $$\Lambda = h^{-1}\arctan[h(M^2-1)^{\frac{1}{2}}] - \arctan[(M^2-1)^{\frac{1}{2}}].$$

In this case, (2.8a) will take the form

(2.12) $$\psi_{\Lambda\Lambda} - \psi_{\theta\theta} + 4N_1\psi_\Lambda = 0, \quad N_1 = \frac{k+1}{8}\frac{M^4}{(M^2-1)^{3/2}}.$$

REMARK. Equations (2.8) and (2.12) can be simplified. If ψ is replaced by

(2.13) $$\psi^* = \psi/R$$

where $(\partial R/\partial \bar{Z}) = N$, then ψ^* satisfies the equations

(2.14) $$\psi^*_{\lambda\lambda} + \psi^*_{\theta\theta} + 4F\psi^* = 0 \text{ and } \psi^*_{\Lambda\Lambda} - \psi^*_{\theta\theta} - 4F_1\psi^* = 0,$$

where

(2.15) $$F = F_1 = \frac{(k+1)M^4}{64}\left[\frac{-(3k-1)M^4 - 4(3-2k)M^2 + 16}{(1-M^2)^3}\right].$$

For subsequent use, we write down the expansions of N and F in the neighborhood of $\lambda = 0$

(2.16) $$N = (1/12\lambda)[1 - \tfrac{1}{4}(k+1)^{\frac{1}{3}}[(2 + \tfrac{3}{5}2^{\frac{1}{3}})k + 5\cdot 2^{\frac{1}{3}} - 2] \times [2^{-1/6}3^{2/3}(k+1)^{-5/6}(-\lambda)^{2/3}] + \cdots],$$

(2.17) $$F = (5/144)(-\lambda)^{-2} + A_{-2}(-\lambda)^{-2/3} + A_0 + A_2(-\lambda)^{2/3} + \cdots = 5/36(-2\lambda)^2 + \cdots.$$

In the vicinity of $\mathrm{H} = 0$, i.e., the sonic line, $l(\mathrm{H})$ may be replaced by the first term in its expansion (2.4). Using this value of $l(\mathrm{H})$ in (2.2),

we obtain the so-called "simplified" compressibility equation (1.1). In considering transonic flows, the solutions of the simplified equation will therefore give a fair approximation—in a certain neighborhood of the sonic line—of the exact streamfunction.

The expression for N, F, and H will, in this case, reduce to

$$
\begin{aligned}
(2.18) \qquad & N = N\dagger = -(1/6)(1/(-2\lambda)), \\
& F = F\dagger = (5/36)(1/(-2\lambda)^2), \\
& \mathrm{H} = \mathrm{H}\dagger = (3^{2/3}/2)(2/(k-1))^{(k-2)/(3k-3)}(-\lambda)^{2/3},
\end{aligned}
$$

respectively.

REMARK. We are using the same variable λ in both the exact and the simplified case, as this facilitates the comparison of the respective flow patterns.

3. Application of integral operators to the compressibility equation. Integral operator of the first kind. The use of integral operators in the theory of the compressibility equation is based on the following theorem:

THEOREM 3.1. *Let $E(Z, \bar{Z}, t)$ be a function of two real and one complex variables, λ, θ, t, which is defined for t along a curve connecting $t = -1$ and $t = 1$, and for $(\lambda, \theta) \varepsilon G$. G denotes here a sufficiently small neighborhood of the origin.*

Let E satisfy the following conditions:

1. *E possesses continuous partial derivatives with respect to all three of its arguments, up to the second order.*

2. *The expression*

$$(3.1) \qquad [(1-t^2)E(Z, \bar{Z}, t)/Zt]\partial/\partial\bar{Z}$$

is continuous for $Z = 0$, and approaches zero, uniformly with respect to $(\lambda, \theta) \varepsilon G$ as $t \to -1$ or $t \to 1$.

3. *E satisfies the partial differential equation*

$$(3.2) \qquad \boldsymbol{G}(E) \equiv (1-t^2)(E_{\bar{Z}t} + \dot{N}E_t) - (1/t)E_{\bar{Z}} + 2tZ\boldsymbol{L}(E) = 0$$

where

$$(3.3) \qquad \boldsymbol{L}(E) \equiv E_{Z\bar{Z}} + N(E_Z + E_{\bar{Z}}) \equiv (1/4)E_{\lambda\lambda} + (1/4)E_{\theta\theta} + NE_\lambda.$$

If $f(\zeta/2)$ is an analytic function of ζ defined in a simply-connected domain P, which includes the origin, then the expression $u(Z, \bar{Z})$, given by

(3.4a) $$u(Z,\bar{Z}) = \mathrm{P}(f),$$

(3.4b) $$\mathrm{P}(f) \equiv \int_{-1}^{1} E(Z,\bar{Z},t) f(\tfrac{1}{2}Z(1-t^2))\,dt/(1-t^2)^{\frac{1}{2}}$$

is defined in a simply-connected domain which lies in $G \cap P$, and satisfies the equation $\boldsymbol{L}(u) = 0$.

The proof of this theorem is given in [**1**, § 1 and **6**, pp. 34-39]. The function f will be called the associate of $\mathrm{P}(f)$ with regard to the operator **P**.

Since (3.2) has an infinity of solutions, there exist infinitely many integral operators; a closer investigation of their properties will show which type of operator is best suited for the purpose on hand. The following property of the integral operator plays an important role in some of the applications.

As is well-known, a harmonic function $G_1(\lambda',\theta') = G(Z',\bar{Z}')$ may be written in the form

$$G(Z',\bar{Z}') = (1/2i)[g(Z') - g(\bar{Z}')].$$

Here [6]

$$Z' = \lambda' + i\theta', \quad \bar{Z}' = \lambda' - i\theta', \quad \lambda' = \lambda - \lambda_0, \quad \theta' = \theta - \theta_0,\ \lambda_0 < 0.$$

If now the harmonic function $G_1(\lambda',\theta')$ is continued to the complex values of the arguments λ' and θ', i. e., if we assume that Z' and $\bar{Z}'$ are not necessarily conjugate to each other, and if, in particular, we consider G and g in the so-called characteristic planes $Z' = 0$ or $\bar{Z}' = 0$, then we see that the analytic function of a complex variable and the continuation of the real harmonic function differ only by constants; [7] indeed,

$$G(Z',0) = (1/2i)[g(Z') - g(0)].$$

An integral operator generates complex solutions of $\boldsymbol{L}$, and we may demand that these complex solutions possess an analogous property. We shall show that there exists an integral operator $\boldsymbol{p}$ (which is connected by relations (3.11), (3.12) with P) such that the complex solution of (3.3)

[6] In the discussions, it will be useful to consider a shift of origin to the point λ_0, θ_0 (point of reference of the operator).

In this section Z' and $\bar{Z}'$ will be treated as independent variables.

[7] Analytic functions of a complex variable represent a very special subclass of complex harmonic functions (i. e., the totality of functions $G + iH$, where G and H are two *arbitrary* real harmonic functions). Operator (3.4a), (3.4b) generates also a subclass of *complex* solutions of the equation $\boldsymbol{L}$.

$$u(Z', \bar{Z}') = \mathbf{p}[g(Z')]$$

and the real solution

$$\psi(Z', \bar{Z}') = \operatorname{Im}[u(Z', \bar{Z}')]$$

are connected by the following relations

$$(3.5) \quad \psi(Z', 0) = (1/2i)[u(Z', 0) - R(Z', 0) \text{ const.}], \quad u(Z', 0) = g(Z')$$

where $R(Z', \bar{Z}')$ is a given function defined in (3.7). This operator P will be called "integral operator of the first kind" and will be denoted by P_1.

Defining the generating function $E_1(Z', \bar{Z}', t)$ of the operator of the first kind by the requirement that

$$(3.6a) \qquad E_1(Z', 0, t) = 1$$

and

$$(3.6b) \qquad E_1(0, \bar{Z}', t) = \exp\left[-\int_0^{\bar{Z}'} N d\bar{Z}'\right],$$

we shall show that these relations imply the property (3.5).

Writing the generating function of the first kind E_1 in the form

$$(3.7) \quad E_1 = R(Z', \bar{Z}') E^*_1(Z', \bar{Z}', t);$$

$$R(Z', \bar{Z}') = \exp\left[-\int_0^{\bar{Z}'} N(Z' + \bar{Z}'_1) d\bar{Z}'_1\right]$$

and assuming that E^*_1 has the development

$$(3.8) \qquad E^*_1 = 1 + \sum_{n=1}^{\infty} Z'^n t^{2n} P^{(n)}(Z', \bar{Z}'),$$

it is found that (3.7) satisfies the relation (3.6b). Substituting this into equation (3.2) we find that the $P^{(n)}$ satisfy the following recurrence relations:

$$(3.9) \quad P_{\bar{Z}'}{}^{(1)} + 2F = 0, \quad (2n+1) P_{\bar{Z}'}{}^{(n+1)} + 2 P_{Z'\bar{Z}'}{}^{(n)} + 2F\, P^{(n)} = 0, \qquad n = 1, 2, 3, \cdots.$$

Finally, (3.6a) is satisfied by imposing upon the $P^{(n)}$ the initial conditions

$$(3.10) \qquad P^{(n)}(Z', 0) = 0, \qquad n = 1, 2, 3, \cdots.$$

By the above requirement, the $P^{(n)}$ and hence the generating functions E_1 (of the first kind) are uniquely determined. Applying the considerations of [**1**, pp. 1173-76] it can be shown that the series (3.8) converges absolutely and uniformly in a sufficiently small neighborhood of the origin $Z' = 0$,

$\bar{Z}' = 0$. The existence of an integral operator of the first kind thus is assured. Assuming that the associate function f is regular in a sufficiently large domain, we prove that by applying to it the integral operator of the first kind, we obtain a solution $u(Z, \bar{Z})$ (see (3.4a), (3.4b)) of (2.8b) defined in a sufficiently small neighborhood of the origin. We shall show in the following that if f is regular for $M < 1$, this solution can be continued throughout the whole subsonic region.

REMARK. Integral operators of the first kind can also be written in a somewhat different form which is useful for various purposes. Namely (as can be shown by a straightforward computation, see [**4**, pp. 618-619]) we have [8]

$$(3.11) \qquad \mathrm{P}(f) = \boldsymbol{p}(g)$$

$$= R(Z', \bar{Z}')[g(Z') + \sum_{n=1}^{\infty} 2^{-2n} \frac{\Gamma(2n+1)}{\Gamma(n+1)} P^{(n)}(Z', \bar{Z}') g^{[n]}(Z')]$$

$$g^{[n]}(Z') = \int_0^{Z'} \int_0^{Z_1} \cdots \int_0^{Z_{n-1}} g(Z_n)\, dZ_n \cdots dZ_1$$

$$= \frac{(-1)^{n-1}}{(n-1)!} \int_{\zeta=0}^{\zeta=Z'} (Z' - \zeta)^{n-1} g(\zeta)\, d\zeta$$

where

$$(3.12) \qquad g(Z') = \int_{t=-1}^{1} f[\tfrac{1}{2}Z'(1-t^2)]\, dt/(1-t^2)^{\frac{1}{2}}.$$

(3.11), (3.7) and (3.10) imply (3.5).

We proceed now to the proof that every *real* solution of equation (2.8b) can be represented in a sufficiently small neighborhood of the origin $Z' = 0$, $\bar{Z}' = 0$ as the imaginary part of the right-hand side of (3.11) with suitably chosen associate function f (or g).

Let $\psi(Z', \bar{Z}')$ ($\bar{Z}'$ being conjugate to Z') be a real solution of equation (2.8b) which is regular in a sufficiently small neighborhood of the origin. Since this equation is of elliptic type and its coefficient N is an analytic function of two variables, $\psi(Z', \bar{Z}')$ can be written in the form of a power series

$$\psi(Z', \bar{Z}') = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} D_{mn} Z'^m \bar{Z}'^n, \qquad D_{mn} = \bar{D}_{nm},$$

which converges in a sufficiently small neighborhood of the origin.

---

[8] f and g are associates of the same solution of (2.8b), the first with respect to the operator P, the second with respect to p. In order to avoid any confusion we shall speak about "p-associate" and "P-associate."

In the plane $\bar{Z}' = 0$, we have:

$$\psi(Z', 0) = \sum_{m=0}^{\infty} D_{m0} Z'^m = G_1(Z') \tag{3.13}$$

and in the plane $Z' = 0$,

$$\psi(0, \bar{Z}') = \sum_{m=0}^{\infty} D_{0m} \bar{Z}'^m = G_2(\bar{Z}')\,; \tag{3.14}$$

G_1 and G_2 are two analytic functions of one complex variable Z' and $\bar{Z}'$, respectively, which are regular in a sufficiently small neighborhood of the origin. (We note that $G_1(0) = G_2(0)$ and $G_2(\bar{Z}') = \overline{G_1(Z')}$ since for $\bar{Z}'$ conjugate to Z', ψ is real.) On the other hand, by classical results (the initial value problem in the theory of partial differential equations), it is known that if functions $G_1(Z')$, $G_2(\bar{Z}')$, $G_1(0) = G_2(0)$, are given, there exists *one and only one* solution $\psi(Z', \bar{Z}')$ of equation (2.8b) such that (3.13) and (3.14) hold. The integral operator (3.4) enables us to write down the solution. Indeed, let us determine two functions say $g_1(Z')$ and $g_2(\bar{Z}')$, $g_2(\bar{Z}') = \overline{g_1(Z')}$, such that

$$g_1(Z') + g_2(0)\overline{R(0, \bar{Z}')} = G_1(Z') \tag{3.15}$$

and therefore:

$$g_2(\bar{Z}') + g_1(0) R(0, \bar{Z}') = G_2(\bar{Z}').$$

Now

$$\begin{aligned} R(Z', \bar{Z}')\Big[g_1(Z') & \\ + \sum_{n=1}^{\infty} \frac{\Gamma(2n+1)}{2^{2n}\Gamma(n+1)} & P^{(n)}(Z', \bar{Z}') \int_0^{Z'} \cdots \int_0^{Z_{n-1}} g_1(Z_n)\, dZ_n \cdots dZ_1\Big] \\ + \bar{R}(\bar{Z}', Z')\Big[g_2(\bar{Z}') & \\ + \sum_{n=1}^{\infty} \frac{\Gamma(2n+1)}{2^{2n}\Gamma(n+1)} & \bar{P}^{(n)}(\bar{Z}', Z') \int_0^{\bar{Z}'} \cdots \int_0^{\bar{Z}_n} g_2(\bar{Z}_n)\, d\bar{Z}_n \cdots d\bar{Z}_1\Big] \end{aligned} \tag{3.16}$$

will represent a solution of (2.8b) which satisfies the conditions (3.13) and (3.14); our assertion that every (real) solution can be represented as the imaginary part of (3.13) is therefore proved.

As already mentioned, (3.18) is a priori only defined in a sufficiently small neighborhood of the origin. We shall prove, however, that provided f is regular for $\lambda' < \lambda_0$, the solution $\psi(Z', \bar{Z}')$ obtained in this way can be continued into the whole region $\mathrm{E}[\mathrm{Re}(Z' + \bar{Z}') < 2\lambda_0, \mid Z' \mid < \infty, \mid \bar{Z}' \mid < \infty]$. As shown in [**6**, pp. 56 ff.; **8**, p. 48 ff.], the quantity $F = F(\frac{1}{2}(Z' + \bar{Z}'))$

introduced in (2.15) is a function of two complex variables Z', $\bar{Z}'$ which is defined in the above region. Therefore the expression [9]

$$(3.17)\quad R(Z',\bar{Z}')[g_1(Z') - \int_0^{Z'}\int_0^{\bar{Z}'} Fg_1 dZ_1 d\bar{Z}_1 + \int_0^{Z'}\int_0^{\bar{Z}'} F[\int_0^{Z_1}\int_0^{\bar{Z}_1} Fg_1 dZ_2 d\bar{Z}_2] dZ_1 d\bar{Z}_1 + \cdots] + \bar{R}(\bar{Z}',Z')[g_2(\bar{Z}') - \int_0^{Z'}\int_0^{\bar{Z}'} Fg_2 dZ'_1 d\bar{Z}'_1 + \int_0^{Z'}\int_0^{\bar{Z}'} [F\int_0^{Z_1}\int_0^{\bar{Z}_1} Fg_2 dZ_2 d\bar{Z}_2] dZ_1 d\bar{Z}_1 + \cdots]$$

satisfies the differential equation (2.8b) and the initial conditions (3.13), (3.14). It is evident that the series (3.17) converges in any simply-connected domain which includes the origin $Z'=0$, $\bar{Z}'=0$ and which is common to the regularity domains of F, g_1, g_2. Since by the above requirements a solution of (2.8b) is uniquely determined, the expressions (3.16) and (3.17) must coincide, so that they are two different representations of the same function.

We proceed now to the discussion of the relations between the domains of regularity, a and k, of the $\boldsymbol{p}_1$-associate function g and the generated solution $\boldsymbol{p}_1(g)$ in the real plane, i.e., for $\bar{Z}'$ conjugate to Z'.

THEOREM 3.2. *Let B be a bounded region of the (real) λ', θ'-plane, situated in* $\mathrm{E}[\lambda' < \lambda_0]$. *If g is regular in B, then $\boldsymbol{p}_1(g)$ is also regular in B; conversely, the regularity of $\boldsymbol{p}_1(g)$ in B implies that g is regular there.*

Proof. In order to prove our statement we investigate the relations which exist between the regularity domain of a solution $\psi(Z',\bar{Z}')$ of (2.8b) in the real λ', θ'-plane (i.e., for $\bar{Z}'$ conjugate to Z') and in the space of two complex variables, λ, θ (i.e., when $\bar{Z}'$ and Z' are two independent complex variables). Let B be a domain in the λ, θ-plane. We denote as the hull [10]

[9] The integral operator of the first kind may be regarded as a generalization of the Riemann formula in the theory of linear hyperbolic equations $u_{XY} + a(X,Y)u_X + b(X,Y)u_Y + c(X,Y) = 0$ to the elliptic case, where the real variables X, Y are replaced by two independent complex variables, Z' and Z' respectively. If we consider the solution in the real plane, i.e., for Z', $\bar{Z}'$ conjugate, we thus obtain solutions of elliptic equations. See [5, pp. 317-318].

[10] The superscript indicates the dimension of the manifold under consideration. In the case where the manifolds are one- or two-dimensional and are situated in the (real) λ, θ-plane, these superscripts are omitted.

$H^4(B)$ of B the intersection $B_1{}^4 \cap B_2{}^4$ of two cylinders of (four-dimensional) space, $B_1{}^4 = \mathrm{E}[Z' \varepsilon B, \bar{Z}' \text{ arbitrary}]$, $B_2{}^4 = \mathrm{E}[Z' \text{ arbitrary}, \bar{Z}' \varepsilon B]$.

It is well-known that a solution $\psi^*(Z', \bar{Z}')$ of the first equation of (2.14) can be represented in the domain B of the real λ, θ-plane in the form

$$(3.18)\quad \psi^*(Z', \bar{Z}') = \int_b \left[\frac{\partial \psi^*(\zeta, \bar{\zeta})}{\partial n_\zeta} \phi^*(Z', \bar{Z}'; \zeta, \bar{\zeta}) - \psi^*(\zeta, \bar{\zeta}) \frac{\partial \phi^*(Z', \bar{Z}'; \zeta, \bar{\zeta})}{\partial n_\zeta}\right]$$

where b denotes the boundary of B, n_ζ the interior normal to b, ds_ζ the line element of b. Here ϕ^* is a fundamental solution of (2.14), and we have

$$(3.19)\quad \phi^* = \tfrac{1}{2}\chi(Z', \bar{Z}'; \zeta, \bar{\zeta})\,[lg(Z' - \zeta) + lg(\bar{Z}' - \bar{\zeta})] + v(Z', \bar{Z}'; \zeta, \bar{\zeta})$$

where [11]

$$(3.20)\quad \chi(Z', \bar{Z}'; \zeta, \bar{\zeta}) = 1 - \int_\zeta^{Z'} \int_{\bar{\zeta}}^{\bar{Z}'} F\,dZ_1 d\bar{Z}_1 + \int_\zeta^{Z'} \int_{\bar{\zeta}}^{\bar{Z}'} F\left[\int_\zeta^{Z_1} \int_{\bar{\zeta}}^{\bar{Z}_1} F\,dZ_2 d\bar{Z}_2\right] dZ_1 d\bar{Z}_1 + \cdots$$

$$(3.21)\quad v(Z', \bar{Z}'; \zeta, \bar{\zeta}) = \int_\zeta^{Z'} \int_{\bar{\zeta}}^{\bar{Z}'} G\,dZ_1 d\bar{Z}_1 - \int_\zeta^{Z'} \int_{\bar{\zeta}}^{\bar{Z}'} F\left[\int_\zeta^{Z_1} \int_{\bar{\zeta}}^{\bar{Z}_1} G\,dZ_2 d\bar{Z}_2\right] dZ_1 d\bar{Z}_1 + \cdots,$$

$$G = -(1/(\bar{Z}' - \bar{\zeta}))(\partial\chi/\partial Z') - (1/(Z' - \zeta))(\partial\chi/\partial \bar{Z}')$$

Since F is defined for all values: $\mathrm{Re}((Z' + \bar{Z}')/2) < \lambda_0$, $|Z| < \infty$, $|\bar{Z}| < \infty$, it follows from (3.17) that if g_1 is regular in B, $B = \mathrm{E}[\lambda' < \lambda_0]$, then $\boldsymbol{p}_1(g_1)$ is regular in $H^4(B)$, and therefore also in the domain B, which is the intersection of $H^4(B)$ with the real λ, θ-plane. This is the first assertion of Theorem 3.2.

From (3.18), (3.19), (3.20), (3.21) it follows that every solution which is regular in the domain B of the real λ, θ-plane can be extended to the complex values in $H^4(B)$. Since B represents the intersection of $H^4(B)$ with the plane $Z' = 0$ as well as with $\bar{Z}' = 0$, $\psi(Z', 0)$ and $\psi(0, \bar{Z}')$ are regular in B. Since $R(Z', \bar{Z}')$ is regular in $\mathrm{E}[\mathrm{Re}((Z' + \bar{Z}')/2) < \lambda_0, |Z'| < \infty, |\bar{Z}'| < \infty]$, it follows from (3.17) that $G_1(Z')$, (as well as $G_2(\bar{Z}') = G_1(Z')$ see (3.15)) is regular in B. See [**1**, § 1 and **2**, § 2]. This completes the proof of Theorem 3.2.

[11] We note that $\chi(Z', \bar{Z}'; 0, 0) = \boldsymbol{p}_1(1)$.

As has been shown in [5, pp. 318-319], the function generated by integral operators $P(f)$ defined in (3.11), will have branchpoints of finite order,[12] at the same points as the function

$$\mathrm{P}_0(f) = \int_{t=-1}^{1} f[\tfrac{1}{2}Z(1-t^2)]\,dt/(1-t^2)^{\frac{1}{2}}. \tag{3.22}$$

As a consequence, the following holds.

THEOREM 3.3. *Suppose that the function*

$$g(Z) = \mathrm{P}_0(f) \tag{3.23}$$

is defined and regular in a region R situated on a Riemann surface, which possesses in its interior a finite number of branchpoints, each of finite order. Let further the projection of R on the schlicht λ, θ-plane lie in $\mathrm{E}[-\infty < \lambda' < \lambda_0]$.

Then the function

$$\mathrm{P}_1(f) = \int_{t=-1}^{1} E_1(Z', \bar{Z}', t) f[\tfrac{1}{2}Z(1-t^2)]\,dt/(1-t^2)^{\frac{1}{2}} \tag{3.24}$$

(see (3.2), (3.7), (3.8), (3.9), (3.10)); *is a solution of* (2.8b) *defined in R, possessing branchpoints at the same points and of the same orders as g.*

$\psi = \mathrm{Im}\,[\mathrm{P}_1(f)]$ *satisfies equation* (2.8a) *and can be interpreted as a streamfunction (in the λ, θ-plane) of a (possible) flow pattern of a compressible fluid.*

Making an obvious modification, we can in a similar way extend the definition of the operator P_1 so that it can be applied to functions of one variable $\Lambda + \theta$ and $\Lambda - \theta$ respectively, thus generating solutions of equation (2.12). The expression

$$\mathrm{P}_1(f_1(\Lambda + \theta)) + \mathrm{P}_1(f_2(\Lambda - \theta)) \tag{3.25}$$

where f_1 and f_2 are two linearly independent functions, will represent a (possible) streamfunction of a supersonic flow pattern.

4. The integral operator of the second kind in the case of the simplified compressibility equation. The integral operator of the first kind, convenient though it is for many purposes, has the disadvantage that it does

[12] At poles and logarithmic singularities, operators $\mathrm{P}(f)$ do not, in general, preserve certain properties of $\mathrm{P}_0(f)$ which are essential in aerodynamical applications; in these cases, we have to use other means (see [3, 9]), in order to produce the necessary singularities of ψ.

not represent solutions of the compressibility equation in the neighborhood of the sonic line. Furthermore, it has the disadvantage from the practical point of view that the $P^{(n)}(Z', \bar{Z}')$ are functions of two variables which makes tabulation of the values of $P^{(n)}$ very time-consuming.

These two disadvantages can be removed by the use of another operator—to be termed "operator of the second kind"—for which the $P^{(n)}$ are functions of one variable only and which yields a representation of the streamfunction in the neighborhood of the sonic line. This operator has a number of other distinctive features which will best be elucidated by the detailed discussion of the so-called "simplified" compressibility equation, i. e., where $N = N\dagger = (12\lambda)^{-1}$ in equation (2. 8a) or alternatively, where $F = F\dagger = 5/144\lambda^2$ in equation (2. 14).

According to Theorem 3. 1, any function E of the form

$$E = HE^*, \tag{4.1}$$

where E^* is a solution of the equation

$$G_2(E^*) = (1 - t^2)E_{\bar{Z}t} - (1/t)E^*_{\bar{Z}} + 2ZtE^*_{Z\bar{Z}} + 2ZtFE^* = 0, \tag{4.2}$$

and H is defined by

$$\begin{aligned} H(2\lambda) &= \exp\left[-\int_{-\infty}^{2\lambda} N(\tau)\,d\tau\right] \\ &= (1 - M^2)^{-1/4}[1 + \tfrac{1}{2}(k-1)M^2]^{-1/2(k-1)} \\ &= S_0(-2\lambda)^{-1/6}p((-2\lambda)^{2/3}) \end{aligned} \tag{4.3}$$

with

$$p((-2\lambda)^{2/3}) = 1 + S_1(-2\lambda)^{2/3} + S_2(-2\lambda)^{4/3} + \cdots,$$

$$S_0 = 2^{(2k+1)/(6k-6)}3^{-1/6}(k+1)^{(2-k)/(6k+6)}$$

$$\begin{aligned} S_1 &= (1/10)(3/4)^{2/3}(k+1)^{-1/3}(2k+5), \\ S_2 &= -(1/1400)(3/4)^{4/3}(k+1)^{-2/3}(64k^2 + 70k + 75), \cdots \end{aligned} \tag{4.4}$$

may be used as a generating function of our operator. In the case of the simplified equation, we have $N\dagger = 1/(12\lambda)$, $p((-2\lambda)^{2/3}) = 1$, $H\dagger(2\lambda) = S_0(-2\lambda)^{-1/6}$, $F\dagger = 5/144\lambda^2$.

We now introduce a new variable

$$u = t^2 Z/(Z + \bar{Z}), \tag{4.5}$$

and we shall show that in the simplified case there exist solutions of (4. 2) which are functions of u alone.

LEMMA 4. 1.

$$(4.6)\quad E^{*}\dagger(\lambda,\theta,t)=A_1\mathrm{F}(1/6,\ 5/6,\ 1/2,\ -t^2(\lambda+i\theta)/-2\lambda) + B_1(-t^2(\lambda+i\theta)/-2\lambda)^{\frac{1}{2}}\mathrm{F}(2/3,\ 4/3,\ 3/2,\ -t^2(\lambda+i\theta)/-2\lambda)$$

(where $\mathrm{F}(\alpha,\ \beta,\ \gamma,\ X)$ denotes the hypergeometric function and A_1 and B_1 are arbitrary constants) is the most general solution of (4. 2) which is a function of u alone. Let us note that there exist other solutions of (4. 2) which are functions of *one* variable, see (4. 20).)

Proof. We shall show that (4. 2) can be reduced to an ordinary differential equation whose solution is (4. 6). A formal computation yields

$$\partial u/\partial t=2u/t,\quad \partial u/\partial Z=(t^2u-u^2)/t^2Z,\quad \partial u/\partial\bar{Z}=-u^2/t^2Z,$$
$$E^{*}\dagger_Z=-u^2t^{-2}Z^{-1}E^{*}\dagger_u,\quad E^{*}\dagger_{Zt}=-2u^2t^{-3}Z^{-1}[uE^{*}\dagger_{uu}+E^{*}\dagger_u]$$
$$E^{*}\dagger_{Z\bar{Z}}=-u^2t^{-4}Z^{-2}[(ut^2-u^2)E^{*}\dagger_{uu}+(t^2-2u)E^{*}\dagger_u],$$
$$F\dagger=(5/36)u^2t^{-4}Z^{-2}.$$

Substituting the above expressions into (4. 2) we obtain

$$G_2(E^{*}\dagger)=-2u^2t^{-3}Z[u(1-u)E^{*}\dagger_{uu}+(\tfrac{1}{2}-2u)E^{*}\dagger_u-(5/36)E^{*}\dagger]=0.$$

The equation

$$(4.7)\qquad u(1-u)E^{*}\dagger_{uu}+(\tfrac{1}{2}-2u)E^{*}\dagger_u-(5/36)E^{*}\dagger=0$$

is a hypergeometric equation whose general solution can be represented in the form

$$(4.8)\quad E^{*}\dagger=A_1\mathrm{F}(1/6,5/6,1/2,u)+B_1u^{\frac{1}{2}}\mathrm{F}(2/3,4/3,3/2,u)\qquad |u|<1$$

$$(4.8')\qquad =A_2u^{-1/6}\mathrm{F}(1/6,2/3,1/3,1/u)+B_2u^{-5/6}\mathrm{F}(5/6,4/3,5/3,1/u),\qquad |u|>1.$$

Replacing u by the right-hand side of (4. 5), we arrive at (4. 6).

Thus, combining (4. 1) and (4. 6), we obtain for the generating function in the simplified case [13]

---

[13] We note that in many instances we may omit the second term on the right-hand side of (4. 9) since

$$\int_{t=-1}^{1}t\mathrm{F}(2/3;\,4/3,3/2,-t^2(\lambda+i\theta)/-2\lambda)f(\tfrac{1}{2}Z(1-t^2))dt/(1-t^2)^{\frac{1}{2}}=0$$

if f is regular at $Z=0$.

$$(4.9)\quad E\dagger(\lambda,\theta,t)=A_1S_0(-2\lambda)^{-1/6}\mathrm{F}(1/6,5/6,1/2,-t^2(\lambda+i\theta)/-2\lambda)$$
$$+B_1S_0(-2\lambda)^{-2/3}[-t^2(\lambda+i\theta)]^{\frac{1}{2}}\mathrm{F}(2/3,4/3,3/2,-t^2(\lambda+i\theta)/-2\lambda),$$
$$|-t^2(\lambda+i\theta)/-2\lambda|<1$$

$$(4.9')\quad =A_2S_0[(-t^2)(\lambda+i\theta)]^{-1/6}\mathrm{F}(1/6,2/3,1/3,-2\lambda/-t^2(\lambda+i\theta)]$$
$$+B_2S_0(-2\lambda)^{-2/3}[(-t^2)(\lambda+i\theta)]^{-5/6}\mathrm{F}(5/6,4/3,5/3,-2\lambda/-t^2(\lambda+i\theta)),$$
$$|-2\lambda/-t^2(\lambda+i\theta)|<1.$$

In Theorem 3.1 we proved the existence of a generating function by means of which we can obtain solutions of the compressibility equation in a sufficiently small neighborhood of the origin. We shall now show that in the case of the simplified equation and the operator of the second type this result in the small can be replaced by a result in the large. This result enables us, from the behavior of the associate f, to make conclusions concerning the behavior of the generated solution of the compressibility equation in its entire domain of definition. An exact formulation of this statement, at first for the subsonic region, is given in the following theorem:

THEOREM 4.1. *Suppose that the function*

$$(4.10)\qquad g(Z)=\int_{t=-1}^{1} f[\tfrac{1}{2}Z(1-t^2)]\,dt/(1-t^2)^{\frac{1}{2}}$$

is regular in a region B (situated on a Riemann surface) which possesses in its interior a finite number of branchpoints,[14] *each of finite order. Let further the projection of B on the schlicht λ,θ-plane lie in* $\mathrm{E}[-\infty<\lambda<0]$.

The function

$$(4.11)\quad \psi=\mathrm{Im}[\mathrm{P}\dagger_2(f)],$$
$$\mathrm{P}\dagger_2(f)=\int_C E\dagger(\lambda,\theta,t)f[\tfrac{1}{2}Z(1-t^2)]\,dt/1-t^2)^{\frac{1}{2}},\qquad Z=\lambda+i\theta$$

(where C is a suitably chosen curve in the complex t-plane connecting the points $t=-1$ and $t=1$) is a solution of (2.8) with $N=N\dagger=1/12\lambda$; this solution is defined in B and possesses branchpoints at the same points and of the same order as (4.10). $\psi=\mathrm{Im}[\mathrm{P}\dagger_2(f)]$ can be interpreted as a streamfunction (in the λ,θ-plane) of a (possible) flow pattern of a compressible fluid (for the simplified compressibility equation).

[14] We assume here that the only singularities of g in B are branchpoints. In applying the integral operator method in the case where g has poles or logarithmic singularities, certain modifications, indicated in [9, p. 469, footnote 14], are needed.

Proof. In order to prove our statement, we have to show that by (4.6) and (4.10) and slight modifications of these formulas, $E\dagger_2(\lambda, \theta, t)$ is defined for all values $\lambda < 0$, $-\infty < \theta < \infty$ and for values t belonging to a suitable simple, sufficiently smooth curve in the complex t-plane which connects $t = -1$ and $t = 1$. Obviously, this curve has to avoid the points $t = 0$, $t = \pm (2\lambda/(\lambda + i\theta))^{\frac{1}{2}}$ as these would give rise to singularities of the hypergeometric function. On the other hand, any such curve will be suitable for our purposes. However, with a view to the subsequent generalization of our procedure to the "exact" case, we shall use two special paths of integration, C_1 and C_2, the former apart from its terminals $t = \pm 1$, inside $\mathrm{E}[\,|\,t\,| < 1]$, and the latter outside $\mathrm{E}[\,|\,t\,| \leq 1]$. C_1 will be used for values λ, θ satisfying $|(\lambda + i\theta)/2\lambda\,| < 1$ and C_2 for the case $|(\lambda + i\theta)/2\lambda\,| > 1$.

It should be noted that the two terminals of the integration path, viz., $t = \pm 1$, will never be singularities of $E\dagger_2(\lambda, \theta, t)$ since, for $t = \pm 1$, $(\lambda + i\theta)/2\lambda \neq 1$ for real λ and θ.

The expressions thus obtained will not necessarily be analytical continuations of each other (*qua* functions of λ, θ). Since, however, the hypergeometric equation has only two linearly independent solutions, the constants, A_1, B_1 and A_2, B_2 can always be so adjusted as to make these two solutions analytical continuations of each other.

REMARK. It would, of course, also be possible to characterize the path of integration in a manner which is topologically invariant with regard to the way the singular points of $E\dagger_2(\lambda, \theta, t)$ are by-passed; using this definition we would, for any value (λ, θ), obtain one and the same function $\psi(\lambda, \theta)$. However, although this procedure has some theoretical advantages, its actual carrying out may give rise to certain difficulties; in practical applications it is much easier to assure analytical continuation by the determination of the constants A_2, B_2 if A_1, B_1 are given, or vice versa.

The integral representation (4.11) can be immediately generalized to the supersonic case where it will produce, in an analogous manner, solutions of (2.12) with $N_1 = N\dagger_1$ (see (2.18)). Indeed, replacing λ by the variable $\omega = \lambda + i\Lambda$ and considering the solution $\psi(\omega, \theta)$ of (2.8a) in the plane $\lambda = 0$, it is seen that $\psi(i\Lambda, \theta)$ satisfies equation (2.12) with $N = 1/12\Lambda$. Repeating the procedure which led to the generating function (4.9) in the subsonic case, we now obtain the generating function

$$(4.12)\qquad E\dagger = (a_1 S_0/(2\Lambda)^{1/6})\mathrm{F}(1/6, 5/6, 1/2, t^2(\Lambda + \theta)/2\Lambda)$$
$$+ (b_1 S_0(\Lambda + \theta)^{\frac{1}{2}} t/(2\Lambda)^{2/3})\mathrm{F}(2/3, 4/3, 3/2, t^2(\Lambda + \theta)/2\Lambda),$$
$$|\,t^2(\Lambda + \theta)/2\Lambda\,| < 1.$$

13

If $F(\alpha_1, \beta_1, \gamma_1, X)$ denotes only the hypergeometric series and not the hypergeometric function, (4.12) has to be replaced, for $|t^2(\Lambda+\theta)/2\Lambda| > 1$, by

$$(4.13)\qquad E\dagger = (a_2 S_0/(t^2(\Lambda+\theta))^{1/6})F(1/6, 2/3, 1/3, 2\Lambda/t^2(\Lambda+\theta)) + (b_2 S_0(2\Lambda)^{2/3}/(t^2(\Lambda+\theta))^{5/6})F(5/6, 4/3, 5/3, 2\Lambda/t^2(\Lambda+\theta)),$$

where the constants a_2, b_2 are easily expressible in terms of a_1, b_1.

If $\theta = \Lambda$ and $t^2 = 1$, there arise certain difficulties, since the hypergeometric functions in (4.12) will then become singular. By the transformation formulas of the hypergeometric function, (4.12) may be written in the neighborhood of $((\Lambda+\theta)/2\Lambda)t^2 = 1$, in the form

$$(4.14)\qquad E\dagger = (a_3 S_0/(2\Lambda)^{1/6})F(1/6, 5/6, 3/2, 1 - t^2(\Lambda+\theta)/2\Lambda) \\ b_3 S_0(2\Lambda/(2\Lambda - t^2(\Lambda+\theta)))^{\frac{1}{2}}F(1/3, -1/3, 1/2, 1 - t^2(\Lambda+\theta)/2\Lambda).$$

In order to avoid the complications which arise from the fact that the second term of (4.14) has a singularity for $\Lambda = \theta$, $t^2 = 1$, we shall therefore take $b_3 = 0$. The function $E\dagger$ will accordingly be of the form

$$(4.15)\qquad E\dagger = (a_3 S_0/(2\Lambda)^{1/6})F(1/6, 5/6, 3/2, 1 - t^2(\Lambda+\theta)/2\Lambda).$$

We note further that all these considerations can be repeated with Λ replaced by $-\Lambda$. Our operator will therefore yield two independent types of solutions, depending on whether the argument of the associate function is taken as $\Lambda+\theta$ or $\Lambda-\theta$.

The exact conditions under which our operator can generate solutions of the compressibility equation in the supersonic case are given in the following theorem:

THEOREM 4.2. *Suppose $f_s(\zeta)$, $s = 1, 2$, are real functions of the real variable ζ and everywhere differentiable with the possible exception of $\zeta = 0$; suppose further that in a fixed neighborhood of $\zeta = 0$, f_s can be approximated to any prescribed degree of accuracy by the expressions of the form $\sum_{n=1}^{N} A_n^{(N)}\zeta^{K_n}$, $K_n \geq 1$. Then*

$$(4.16)\qquad \psi(\Lambda,\theta) = \boldsymbol{R}\dagger_1(f_1) + \boldsymbol{R}\dagger_2(f_2),$$

$$\boldsymbol{R}\dagger_s(f_s) = \int_{t=-1}^{1} E\dagger_2(\Lambda, -(-1)^s\theta, t) f_s[\tfrac{1}{2}(\Lambda - (-1)^s\theta)(1-t^2)]dt/(1-t^2)^{\frac{1}{2}}, \qquad s = 1, 2$$

represents a solution of the compressibility equation, which is defined for any $\Lambda > 0$ and can be interpreted as a streamfunction of a (possible) supersonic flow pattern.

Proof. By (4.12) we have

(4.17) $\boldsymbol{R}\dagger_1(f_1) = a_1/(2\Lambda)^{1/6}$

$$\times \int_{-1}^{1} \mathrm{F}(1/6, 5/6, 1/2, t^2((\Lambda+\theta)/2\Lambda) f((\Lambda+\theta)(1-t^2)/2)(1-t^2)^{-\frac{1}{2}} dt$$

$$+ c_1(\Lambda+\theta)^{\frac{1}{2}}(2\Lambda)^{-2/3}$$

$$\times \int_{-1}^{1} t\mathrm{F}(2/3, 4/3, 3/2,\ t^2(\Lambda+\theta)/2\Lambda) f((\Lambda+\theta)(1-t^2)/2)(1-t^2)^{-\frac{1}{2}} dt.$$

In view of $\Lambda > 0$, the only values of (Λ, θ) for which this expression may not be differentiable are those for which $\Lambda + \theta = 0$ or for which $u = t^2(\Lambda+\theta)/2\Lambda$ coincides with either of the values $0, 1, \infty$ for $-1 < t < 1$. The case $u = \infty$ is ruled out because of $\Lambda > 0$; the case $u = 1$, although corresponding to a singularity of the hypergeometric equation, does not give rise to a singularity of $\boldsymbol{R}$, by virtue of our particular choice (4.15) of $E\dagger$. The case $u = 0$ need only to be considered for $\Lambda + \theta = 0$, as $t = 0$ obviously does not cause any difficulties.

Under our assumptions, it is sufficient to consider the special case

(4.18) $$f_1(\zeta) = f_{1,\kappa}(\zeta) = 2^{\kappa}\zeta^{\kappa}, \qquad \kappa \geq 1,$$

in which (4.17) reduces to

(4.19) $\boldsymbol{R}\dagger_{2,1}(f_{1,\kappa}) = a_1/(2\Lambda)^{1/6}$

$$\times \int_{-1}^{1} \mathrm{F}(1/6, 5/6, 1/2,\ t^2(\Lambda+\theta)/2\Lambda)(\Lambda+\theta)^{\kappa}(1-t^2)^{\kappa-\frac{1}{2}} dt$$

$$+ c_1(2\Lambda)^{-2/3} \int_{-1}^{1} t\mathrm{F}(2/3, 4/3, 3/2,\ t^2(\Lambda+\theta)/2\Lambda)(\Lambda+\theta)^{\kappa+\frac{1}{2}}(1-t^2)^{k-\frac{1}{2}} dt.$$

For $\Lambda + \theta \to 0$, this expression is obviously continuous. The same is also true of the derivative $\partial\boldsymbol{R}\dagger_{2,1}/\partial \mathrm{H}\dagger$. Indeed, $\partial\boldsymbol{R}\dagger_{2,1}/\partial \mathrm{H}\dagger$ (see (2.18)) behaves in the neighborhood of $\Lambda + \theta = 0$ like $(\Lambda+\theta)^{\kappa-1}$; in view of $\kappa \geq 1$, it therefore remains continuous there. This completes the proof of Theorem 4.1.

Before we proceed to investigate the behavior of our solutions on the sonic line, we note that (4.6) is not the only solution which depends only on one variable. If we write

(4.20) $$\mu = t^2(\lambda - \lambda_0 + i\theta)/2\lambda$$

and try to solve (4.2) by a function of one variable μ, a formal computation shows that $E^*\dagger$ as a function of μ must satisfy the hypergeometric equation (4.7) with u replaced by μ.

Accordingly, the generating function (4.6) for the operator of the second kind may be replaced by

(4.21) $E^{\dagger}(\lambda, \theta, t) = A_1(-2\lambda)^{1/6}\mathrm{F}(1/6, 5/6, 1/2, t^2(\lambda - \lambda_0 + i\theta)/2\lambda)$.

Besides its greater generality, the generating function (4.21 with $\lambda_0 \neq 0$ has a number of additional features, which make it superior to (4.6) in many cases. In (4.6) the point $\lambda = 0$, $\theta = 0$ is a singularity since, by approaching this point in a suitable manner, the argument of the hypergeometric function can be given an arbitrary value. In (4.21) such a singularity does not exist, since λ and $\lambda - \lambda_0 + i\theta$ cannot vanish simultaneously if $\lambda_0 \neq 0$.

Another singularity, which can be removed by using (4.21) with $\lambda_0 \neq 0$ instead of (4.6), occurs in the supersonic case. For $t^2 = 1$ and $\Lambda = \theta$, the second term of the right-hand side of (4.14) becomes singular. In order to allow for this case, we had to assume that $b_3 = 0$, thus somewhat restricting the generality of the solutions we could obtain. Setting $\lambda = i\Lambda$, it is seen that in the supersonic case the argument of the hypergeometric function in (4.21) becomes

(4.22) $$\mu = t^2(\Lambda + i\lambda_0 + \theta)/2\Lambda.$$

For $t = \pm 1$, we have $\mu \neq 1$ for real values of Λ, θ; the singularity in question is therefore removed and the constant b_3 in (4.14) may now take any arbitrary value. Moreover, (4.22) shows that $\mu \neq 0$ for $\Lambda + \theta = 0$; as a consequence, the line $\Lambda + \theta = 0$ loses its singular character and the discussion of the equivalent of (4.17) is considerably simplified.

Thus, by the use of integral operators of the second kind with the generating functions (4.9) and (4.12) as well as (4.21) we obtain solutions $\psi = \mathrm{Im}[\boldsymbol{p}_2(g)]$ which are defined in the subsonic and supersonic regions respectively. If g is defined in a (not necessarily schlicht) domain G_1 in the subsonic region and has, as its only singularities in G_1, branchpoints of finite order (but not poles or logarithmic singularities), then the generated function is again defined in G_1. (In particular, it has branchpoints at the same points and of the same order as g).

If g is a twice differentiable function of one real variable in a schlicht domain G_2 in the supersonic region, the generated function will be a solution of (2.18).

Thus by this procedure, we can generate solutions of (2.2) which are defined in adjacent domains, one in the subsonic, the other in the supersonic region. Our object is to show that these solutions can be continued across the sonic line.

This fact is immediately seen if we introduce a new variable,[15]

$$(4.23)\qquad \begin{aligned} s &= (-\lambda)^{2/3} \quad \text{for} \quad \lambda < 0, \\ s &= -\Lambda^{2/3} \quad \text{for} \quad \Lambda > 0. \end{aligned}$$

The variable $s = s(M\dagger)$ considered as a function of Mach number $M\dagger$, possesses the property that $s(1) = 0$, and that

$$(4.24)\qquad ds(M\dagger)/dM\dagger = -2^{5/3}(3k+3)^{-2/3}M\dagger + O(1 - M\dagger^2)$$

is non-vanishing and bounded in a sufficiently small neighborhood of $M\dagger = 1$.

If $\lambda_0 \neq 0$, and if g is regular for $(-\theta_0 \leq \theta \leq \theta_0,\ s = \lambda = \Lambda = 0)$ then the generated function is an analytic function of θ and s (and therefore of θ and $M\dagger$) for $(-\theta_0 \leq \theta \leq \theta_0,\ -s^{(0)} \leq s \leq s^{(0)})$, $s^{(0)}$ sufficiently small.

If $\lambda_0 = 0$, we have to assume that $\lim_{Z=0} Z^{-5/6} f(Z)$ exists (or alternately that $\lim_{Z=0} Z^{1/6} g'(Z)$ exists) in order to assume that the generated function $\psi(\theta, s)$ is regular also at the point $\theta = 0$, $s = 0$.

5. Integral operator of the second kind in the case of the "exact" compressibility equation. As we mentioned before, any solution E^* of the equation

$$(5.1)\qquad E^*_{\bar{Z}t} - t^2 E^*_{\bar{Z}t} - (1/t)E^*_{\bar{Z}} + 2ZtE^*_{\bar{Z}Z} + 2ZtFE^* = 0,$$

multiplied by $\exp\left[-\int_{-\infty}^{Z+\bar{Z}} N d\tau\right]$, (see (4.3)), is a generating function of an integral operator $\mathrm{P}(f)$ (see (3.4b) which produces solutions of equation (2.8b). The series $1 + \sum_{n=1}^{\infty} (t^2 Z)^n Q^{(n)}(2\lambda)$, where the $Q^{(n)}$'s are solutions of the system:

$$(5.2)\qquad (2n+1)Q_\lambda^{(n+1)} + Q_{\lambda\lambda}^{(n)} + 4FQ^{(n)} = 0,\ Q^{(1)} = -4\int_{-\infty}^{2\lambda} F d\lambda$$

(see (87) of [**6**]), is a solution of (5.1). The above series converges for $|Z| < 2|\lambda|$ and reduces to the first summand of (4.6) for $F = F\dagger$.

It is therefore desirable to obtain solutions of (5.1) which are defined in $|Z| > 2|\lambda|$. In this section we shall determine two power series $E^{*(\kappa)}$ $(\kappa = 1, 2)$, both converging in $|Z| > 2|\lambda|$, such that in the simplified case, $A_2 RE^{*(1)} + B_2 RE^{*(2)}$ reduces to (4.9′).

---

[15] We note that in the simplified case we have

$$s = -2 \cdot 3^{-2/3}(2/(k-1))^{(2-k)/(3k-3)}\mathrm{H}\dagger.$$

THEOREM 5.1. *Let*

$$q^{(n,\kappa)} \equiv \sum_{\nu=0}^{\infty} C_\nu^{(n,\kappa)} (-\lambda)^{n-\frac{1}{2}+(2/3)(\kappa+\nu)} \qquad \kappa = 1, 2, \tag{5.3}$$

be a set of functions which are connected by the relations

$$q_{\lambda\lambda}^{(0,\kappa)} + 4Fq^{(0,\kappa)} = 0 \tag{5.4a}$$

$$2(n + \tfrac{2}{3}\kappa) q_\lambda^{(n,\kappa)} + q_{\lambda\lambda}^{(n+1,\kappa)} + 4Fq^{(n+1,\kappa)} = 0, \qquad n = 1, 2, \cdots, \ \kappa = 1, 2. \tag{5.4b}$$

Then each of the functions

$$E^{*(\kappa)} = \sum_{n=0}^{\infty} q^{(n,\kappa)} / (-t^2 Z)^{n-\frac{1}{2}+(2/3)\kappa} \tag{5.5}$$

is a solution of (5.1). *Each of the series converges in* $E[2\,|\lambda| < |Z|\,]$.

Proof. Substituting the series (5.5) into (5.1) and equating the coefficients of $t^{-(6n+4)/3}(-Z)^{-(6n+1)/6}$, $n = -1, 0, 1, \cdots$, to 0, we obtain the following set of equations:

$$q_{\bar{Z}Z}^{(0,\kappa)} + Fq^{(0,\kappa)} = 0, \ (n + \tfrac{2}{3}\kappa) q_{\bar{Z}}^{(n,\kappa)} + q_{\bar{Z}Z}^{(n+1,\kappa)} + Fq^{(n+1,\kappa)} = 0, \qquad n = 0, 1, \cdots, \tag{5.6}$$

which, if we assume that the $q^{(n,\kappa)}$ are functions of λ alone, result in the equations (5.4a) and (5.4b).

According to (2.17), F can be written in the form

$$F = s^{-3} S(s), \qquad s = (-\lambda)^{2/3}, \tag{5.7}$$

where $S(s) = \alpha_0 + \alpha_1 s + \alpha_2 s^2 + \cdots$, $\alpha_0 = 5/144$, $\alpha_1 = 0, \cdots$, see (2.17), considered as a function of s, is regular in a circle of radius s_0 say, with center at the origin.

Introducing the variable s, we obtain

$$\begin{aligned} q_\lambda^{(n,\kappa)} &= -(ds/d(-\lambda))(dq^{(n,\kappa)}/ds) = -\tfrac{2}{3}(-\lambda)^{-1/3}(dq^{(n,\kappa)}/ds) \\ &= -\tfrac{2}{3}s^{-\frac{1}{2}}(dq^{(n,\kappa)}/ds) \\ q_{\lambda\lambda}^{(n+1,\kappa)} &= (4/9)[-\tfrac{1}{2}s^{-2}q_s^{(n+1,\kappa)} + s^{-1}q_{ss}^{(n+1,\kappa)}]. \end{aligned}$$

Hence, the system (5.4a), (5.4b) assumes the form

$$s^2 q_{ss}^{(0,\kappa)} - \tfrac{1}{2}s q_s^{(0,\kappa)} + 9S(s) q^{(0,\kappa)} = 0 \tag{5.9a}$$

$$-3(n + \tfrac{2}{3}\kappa) s^{5/2} q_s^{(n,\kappa)} + s^2 q_{ss}^{(n+1,\kappa)} - \tfrac{1}{2}s q_s^{(n+1,\kappa)} + 9S(s) q^{(n+1,\kappa)} = 0, \qquad n = 0, 1, 2, 3, \cdots. \tag{5.9b}$$

LEMMA 5.1. There exists a system of solutions $q^{(n,\kappa)}$, $n=0,1,2,\cdots$, of (5.9a), (5.9b) of the form

(5.10) $$q^{(n,\kappa)}=s^{(3/2)(n-\frac{1}{3}+(2/3)\kappa)}T_\kappa^{(n)}(s),\ T_\kappa^{(n)}(s)=\sum_{\nu=0}^{\infty}C_\nu^{(n,\kappa)}s^\nu,\ C_0^{(n,\kappa)}\neq 0$$

where each $T_\kappa^{(n)}(s)$ is a power series which converges in the circle $|s|<s_0$.

Proof. We shall first prove the above lemma for $q^{(0\kappa)}$ and then for arbitrary n by induction on n, i.e., we shall show that if the lemma holds for $q^{(n,\kappa)}$, then it must hold for $q^{(n+1,\kappa)}$. Let us first consider the homogeneous equation

(5.11) $$s^2w_{ss}-\tfrac{1}{2}sw_s+9S(s)w=0,$$

The indicial equation (see [15], p. 225) is

(5.12) $$\rho(\rho-1)-\tfrac{1}{2}\rho+5/16=0,\ \text{or}\ \rho_1=5/4,\ \rho_2=1/4.$$

By substituting $w=s^{1/4}y$, we obtain the equation

(5.13) $$y''+9s^{-2}[S(s)-5/144]y=0.$$

Since $s^{-2}[S(s)-5/144]$ is a regular function of s for $|s|<s_0$, (see 5.7)) we may choose the two particular solutions of (5.13) as

$$y_1=C_0^{(0,1)}+C_2^{(0,1)}s^2+C_3^{(0,1)}s^3+\cdots$$
$$y_2=C_0^{(0,2)}s+C_1^{(0,2)}s^2+C_2^{(0,2)}s^3+\cdots$$

which yield

$$q^{(0,1)}=C_0^{(0,1)}s^{1/4}+C_2^{(0,1)}s^{9/4}+C_3^{(0,1)}s^{13/4}+\cdots\equiv s^{1/4}W^{(1)}(s)$$
$$q^{(0,2)}=C_0^{(0,2)}s^{5/4}+C_1^{(0,2)}s^{9/4}+C_2^{(0,1)}s^{13/4}+\cdots\equiv s^{5/4}W^{(2)}(s).$$

In the case of the simplified equation we have [see (4.9′), (4.1), (4.3)]

$$q\dagger^{(0,1)}=C\dagger_0^{(0,1)}(-\lambda)^{1/6},\ C\dagger_0^{(0,1)}=2^{1/6},$$
$$q\dagger^{(0,2)}=C\dagger_0^{(0,2)}(-\lambda)^{5/6},\ C\dagger^{(0,2)}=2^{5/6}.$$

In order to obtain analogous series for the "exact" case we choose

(5.14) $$C_0^{(0,1)}=C\dagger_0^{(0,1)}=2^{1/6},\qquad C_0^{(0,2)}=C\dagger_0^{(0,2)}=2^{5/6}.$$

Let us now instead of (5.9a) consider the solution of the non-homogeneous equation (5.9b) and let us assume that for $n\geq 0$, we have already proved that $q^{(n,\kappa)}$ has the form (5.10) and $T_\kappa^{(n)}(s)$ converges for $|s|<s_0$.

In order to prove that there exists a solution of the equation (5.9b) of the form $q^{(n+1,\kappa)}=s^{(3/2)(n+\frac{2}{3}+(2/3)\kappa)}T_\kappa^{(n+1)}(s)$, we proceed as follows:

Let

(5. 15) $$q^{(n+1,\kappa)}(s) = w(s)u(s),$$

where $w(s) = s^{\kappa-3/4}W^{(\kappa)}(s)$ is a solution (see above) of the homogeneous equation (5. 11). Then u will satisfy the equation

(5. 16) $$wu_{ss} + (2w_s - w/2s)u_s = 3(n + \tfrac{2}{3}\kappa)s^{\frac{1}{2}}q_s^{(n,\kappa)}.$$

The particular solution of this equation is easily verified to be

(5. 17) $$u = 3(n + \tfrac{2}{3}\kappa)\int_0^s w^{-2}s^{\frac{1}{2}}\left(\int_0^s wq_s^{(n,\kappa)}ds\right)ds$$

Expanding $q^{(n+1,\kappa)}$, we obtain

(5. 19) $$q^{(n+1,\kappa)}(s) = s^{(3/2)(n+\frac{1}{2}+(2/3)\kappa)}T_\kappa^{(n+1)}(s),$$

$$T_\kappa^{(n+1)}(0)$$
$$= [(6n - 3 + 4\kappa)(3n + 2\kappa)/(3n + 3)(3n + 4\kappa - 3)]T_\kappa^{(n)}(0) \neq 0$$

the series for $T_\kappa^{(n+1)}(s)$ converging for $|s| < s_0$.

This completes the proof of Lemma 5. 1. In order to prove Theorem 5. 1, it remains to be shown that it is legitimate to interchange the order of summation in the double series

(5. 20) $$\sum_{n=0}^{\infty}(-t^2Z)^{-(n-\frac{1}{2}+(2/3)\kappa)}\sum_{\nu=0}^{\infty}C_\nu^{(n,\kappa)}(-\lambda)^{n-\frac{1}{2}+(2/3)(\kappa+\nu)}$$

for $2|\lambda| < |Z|$. For this purpose, we shall prove [16]

LEMMA 5. 2.

(5. 21) $$|C_\mu^{(n+1,1)}| \leqq 2^{n+1}\mathrm{M}/s_1^{\mu}$$

for $n + \frac{2}{3}\mu \geqq p_0$, where $C_\mu^{(n+1,1)}$ are the coefficients of the series (5. 2). Here, p_0 and M are sufficiently large constants, and $s_1 = s_0(1+\epsilon)^{-1}$.

Proof. We shall give a proof by induction. Consider at first the $C_0^{(n,1)}$, $n = 0,1,2,\cdots$. If we substitute the power series (5. 3) into (5. 4b), it becomes evident that the $C_0^{(n,1)}$ depend only upon α_0, and are independent of the remaining coefficients α_n, $n > 0$, of the series expansion of $S(s)$, (see (5. 7)). On the other hand, if we substitute $\alpha_0 = 5/144$, $\alpha_n = 0$ for $n > 0$, we obtain the simplified case considered in (4. 9′), where we had a representation for $E\dagger$ as a power series in $(2\lambda/-t^2Z)$. Using the fact that

[16] We shall consider here only the case when $\kappa = 1$. Exactly the same proof holds for $\kappa = 2$.

$E^*\dagger = E\dagger/H\dagger = S_0^{-1}(-2\lambda)^{1/6}E\dagger$ (see (4. 1) ff.) we obtain two power series for $E^*\dagger$; the first of which corresponds to the case $\kappa = 1$ and the second to the case $\kappa = 2$. Thus, setting $A_2 = 1$ in (4. 9′), we have

$$(5.22) \qquad E^*\dagger = (-2\lambda/-t^2Z)^{1/6}F(1/6, 2/3, 1/3, -2\lambda/-t^2Z)$$
$$= (-2\lambda/-t^2Z)^{1/6}$$
$$+\sum_{n=1}^{\infty}\frac{\frac{1}{6}(\frac{1}{6}+1)\cdots(\frac{1}{6}+n)\cdot\frac{2}{3}(\frac{2}{3}+1)(\frac{2}{3}+n)}{\frac{1}{3}(\frac{1}{3}+1)\cdots(\frac{1}{3}+n)\cdot n!}(-2\lambda/-t^2Z)^{n+1/6}$$

whence

$$(5.23) \quad C_0^{(0,1)} = C\dagger_0^{(0,1)} = 2^{1/6}, \; C_0^{(n,1)} = C\dagger_0^{(n,1)}$$
$$= \frac{\frac{1}{6}(\frac{1}{6}+1)\cdots(\frac{1}{6}+n)(\frac{2}{3})(\frac{2}{3}+1)\cdots(\frac{2}{3}+n)2^{n+1/6}}{\frac{1}{3}(\frac{1}{3}+1)\cdots(\frac{1}{3}+n)\cdot n!}, \qquad n = 1, 2, 3, \cdots,$$

from which the inequality (5. 21) for $\mu = 0$ follows.

REMARK. Since in this case $\nu = \mu = 0$, the number s_0 in (5. 21) where ν is replaced by μ) may be given any positive value. For $n = -1$, and an arbitrary μ, the inequality (5. 21) follows from the fact that $q^{(0,1)} = \sum_{\mu=0}^{\infty} C_\mu^{(0,1)} s^{\mu+1/4}$ is a solution of (5. 9a) and the fact that $S(s) = \sum_{\mu=0}^{\infty}\alpha_\mu s^\mu$ is regular in the circles of the radius s_0 (see also (5. 13)); accordingly,

$$(5.24) \qquad |\alpha_\mu| \leq \Gamma/s_0^\mu$$

holds, if Γ is a sufficiently large constant.

We now proceed to the proof (by induction) of (5. 21) for $n > -1$ and $\mu > 0$. Let us assume that this inequality holds for some $n+1$, and $\mu \leq \nu - 1$ as well as for $N \leq n$ and $\mu \leq \nu + 1$. We shall prove that (5. 21) then holds for $N = n + 1$, $\mu = \nu$.

If we substitute the series (5. 3) (with $\kappa = 1$) into (5. 4b), we obtain the relation

$$(5.25) \quad \begin{aligned} &-2(n+\tfrac{2}{3})(n+\tfrac{1}{6}+\tfrac{2}{3}\nu)C_\nu^{(n,1)} + (n+\tfrac{7}{6}+\tfrac{2}{3}\nu)(n+\tfrac{1}{6}+\tfrac{2}{3}\nu)C_\nu^{(n+1,1)} \\ &+ 4\alpha_0 C_\nu^{(n+1,1)} + \sum_{\mu=0}^{\nu-1} 4C_\mu^{(n+1,1)}\alpha_{\nu-\mu} = 0. \end{aligned}$$

Hence,

$$(5.26) \quad \begin{aligned} &|C_\nu^{(n+1,1)}[1 + (4\alpha_0/(n+\tfrac{7}{6}+\tfrac{2}{3}\nu)(n+\tfrac{1}{6}+\tfrac{2}{3}\nu))]| \\ &\leq 2^{n+1}\mathrm{M}/s_0 s_1^{\nu-1}\{[(1+2/3n)/(1+7/6n+(2/3)(\nu/n))] \\ &+ 4\nu\Gamma/(n+\tfrac{7}{6}+\tfrac{2}{3}\nu)(n+\tfrac{1}{6}+\tfrac{2}{3}\nu)\}. \end{aligned}$$

If $n + \frac{2}{3}\nu > p_0$, then

$$(5.27)\qquad | C_\nu^{(n+1,1)} [1 + 4\alpha_0/p_0^2] | \leqq 2^{n+1}\mathrm{M}/s_0 s_1^{\nu-1}[1 + 6\Gamma/p_0],$$

or

$$| C_\nu^{(n+1,1)} | \leqq 2^{n+1}\mathrm{M}(1+\epsilon)/s_0 s_1^{\nu-1} = 2^{n+1}\mathrm{M}/s_1^{\nu}.$$

This completes the proof of Lemma 5.2.

In order to show that it is legitimate to interchange the order of summation in (5.20), we shall show that for

$$| \lambda |^{2/3} \leqq s_0 - \epsilon, \; | -2\lambda/-t^2 Z | \leq 1-\epsilon, \text{ where } \epsilon > 0,$$

the series converges absolutely and uniformly. Indeed, by Lemma 5.2 we obtain

$$(5.28)\qquad \Big| \sum_{n=0}^{\infty} (t^2 Z)^{-(1/6)-n} \sum_{\nu=0}^{\infty} {}_\nu^{(n,1)} (-\lambda)^{n+(1/6)+(2/3)\nu} \Big|$$
$$\leqq \mathrm{M} \sum_{n=0}^{\infty} \sum_{\nu=0}^{\infty} 2^n (-\lambda)^{n+(1/6)+(2/3)\nu}/s_1^{\nu} \, | t^2 Z |^{(1/6)+n}$$
$$= \mathrm{M} \, | -2\lambda/-t^2 Z |^{1/6} [1 - | -2\lambda/-t^2 Z |]^{-1} [1 - | \lambda^{2/3}/s_1 |]^{-1}$$

which for $|\lambda|^{2/3} < s_0 - \epsilon$, $|-2\lambda/-t^2 Z| < 1-\epsilon$, $0 < \epsilon < 1$, becomes smaller than $\mathrm{M}(1-\epsilon)^{1/6}/\epsilon^2$, which shows that the series converges absolutely and uniformly.

In order to obtain a continuation of a given streamfunction to the supersonic region, i.e., for imaginary values of λ, it is convenient to replace λ by the variable $s = (-\lambda)^{2/3}$ used before in a different context (see (4.23)). Then the generating function in the subsonic case may be written as

$$(5.29)\qquad E^{(\kappa)}(\lambda, \theta, t) = H(2\lambda) E^{*(\kappa)}(\lambda, \theta, t)$$
$$= S_0 2^{-1/6} p(s) s^{\kappa-1} \sum_{n=0}^{\infty} \sum_{\nu=0}^{\infty} C_\nu^{(n,\kappa)} s^{(3/2)n} s^{\nu} s^{\nu} / (-t^2 Z)^{n-\frac{1}{3}+(2/3)\kappa},$$
$$| s | < s_0, \; 2 | s |^{3/2} < | Z |, \; \kappa = 1, 2,$$

(see (4.3) and (5.5)) where now

$$(5.29\text{a})\qquad Z = -s^{3/2} + i\theta.$$

By replacing λ by $i\Lambda$, we see that $(-\lambda)^{2/3}$ is changed to $-\Lambda^{2/3}$. Thus we obtain as the generating function in the supersonic case

$$(5.30)\qquad E^{(\kappa)}(\Lambda, \theta, t)$$
$$= 2^{-1/6} S_0 p(s) s^{\kappa-1} \sum_{n=0}^{\infty} \sum_{\nu=0}^{\infty} i^n C_\nu^{(n,\kappa)} (-s)^{(3/2)n} s^{\nu} / (-t^2 Z)^{n-\frac{1}{3}+(2/3)\kappa}$$
$$| s | < s_0, \; 2 | s |^{3/2} < | Z |,$$

where

(5. 30a) $$Z = i(-s)^{3/2} + i\theta.$$

Let us now restrict ourselves to a neighborhood of the point $\lambda = 0$, $\theta = \theta_0$ ($\theta_0 \neq 0$) lying entirely in the domain $D = \mathrm{E}[|s| < s_0, |s| < 3^{-1/3}|\theta|^{2/3}]$.

For $(s, \theta) \varepsilon \mathrm{E}[s^2 + (\theta - \theta_0)^2 < \theta_0^2/25]$, we have

(5. 31a) $$Z^\gamma = (i\theta_0)^\gamma[1 - (s^{3/2} - i(\theta - \theta_0))/i\theta_0]^\gamma = \sum_{m,n} a_{mn}(\theta - \theta_0)^m s^{(3/2)n} \qquad \text{for } M < 1,$$

(5. 31b) $$Z^\gamma = (i\theta_0)^\gamma[1 + ((-s)^{3/2} + (\theta - \theta_0))/\theta]^\gamma = \sum_{m,n} b_{mn}(\theta - \theta_0)^m s^{(3/2)n} \qquad \text{for } M > 1$$

because then both $|(s^{3/2} - i(\theta - \theta_0))/i\theta_0|$ and $|((-s)^{3/2} + i(\theta - \theta_0))/\theta_0|$ are less than

$$(|s|^{3/2} + |\theta - \theta_0|)/|\theta_0| < (3^{-1/2}|\theta| + |\theta - \theta_0|)/|\theta_0| < (2 \cdot 3^{1/2} + 1)/5 < 1.$$

Here a_{mn} and b_{mn} are suitably chosen constants. Thus both the subsonic generating function and the associate function, and therefore $\mathrm{P}[f(-s^{3/2} + i\theta)]$, may be expanded in integral powers of $(\theta - \theta_0)$ and $s^{1/2}$. Similarly, in the supersonic case, $\mathrm{P}[f(i(-s)^{3/2} + i\theta)]$ may be expanded in integral powers of $(\theta - \theta_0)$ and $(-s)^{1/2}$. Both functions are determined and equal to each other for $s = 0$.

We shall show that if f is regular in D, the coefficients of $s^{\nu/2}$, ν odd, vanish, and therefore $\mathrm{P}[f(-s^{3/2} + i\theta)]$ is an analytic function of s and $(\theta - \theta_0)$ at $(0, \theta_0)$.

THEOREM 5. 2. *If f is regular in the domain $D = \mathrm{E}[|s| < s_0, |s| < 3^{-1/3}|\theta|^{2/3}]$, then the solution $\mathrm{P}[f(-s^{3/2} + i\theta)]$ is a regular function of s and $(\theta - \theta_0)$ in D.*

Proof. We have shown that $\mathrm{P}(f)$ may be expresed as a power series in $s^{1/2}$ of the form

(5. 32) $$\mathrm{P}(f) = \sum_{n=0}^{\infty} h_n(\theta - \theta_0) s^{n/2}.$$

Using the variable s, and noting the fact that $N = (1/12)s^{-3/2}(1 + \sum_{n=0}^{\infty} a_{n+1}s^{n+1})$ (see (2. 16)) equation (2. 8a) becomes

(2. 8′) $$4\psi_{ss} + 9s\psi_{\theta\theta} + 2\psi_s \sum_{n=0}^{\infty} a_{n+1}s^n = 0.$$

By substituting (5. 32) in (2. 8′) we obtain the recursion formulae:

(5. 33a) $$h_1 = h_3 = 0$$

(5.33b) h_{n+4}

$$= -[1/(n+4)(n+2)][9h''_{n-2} + 2\sum_{j=0}^{(n-1)/2}((3/2)+j)a_{(n/2)+\frac{1}{2}-j}h_{2j+3}]$$
for n odd,

(5.33c) h_{n+4}

$$= -[1/(n+4)(n+2)][9h''_{n-2} + 2\sum_{j=0}^{n/2}(1+j)a_{(n/2)+1-j}h_{2j+2}]$$
for n even.

Since h_n, where n is odd, depends only on the previous h_m for odd m, we see from (5.33a) that $h_n = 0$ for all odd n.

Thus (5.32) becomes

$$\mathrm{P}[f(-s^{3/2}+i\theta)] = \sum_{n=0}^{\infty} f_n(\theta-\theta_0)s^n. \tag{5.32'}$$

Since s is unchanged by the result of putting $i\Lambda$ for λ, we obtain

$$\mathrm{P}[f(i(-s)^{3/2}+i\theta)] = \sum_{n=0}^{\infty} f_n(\theta-\theta_0)s^n. \tag{5.34}$$

The expressions $\mathrm{P}[f(-s^{3/2}+i\theta)]$ and $\mathrm{P}[f(i(-s)^{3/2}+i\theta)]$, *qua* functions of s and $(\theta-\theta_0)$, are analytic continuations of each other across the sonic line.

Thus, assuming that the associate function is regular in a sufficiently large domain, and applying the integral operator of the second kind, we obtain solutions of a compressibility equation defined in four adjacent domains,

$$D_1 = \mathrm{E}[M<1, \theta>3^{1/2}\,|\lambda(M)|] + \mathrm{E}[M>1, \theta>\Lambda(M)],$$
$$D_2 = \mathrm{E}[M<1, |\theta|<3^{1/2}\,|\lambda(M)|],$$
$$D_3 = \mathrm{E}[M<1, \theta<-3^{1/2}\,|\lambda(M)|] + \mathrm{E}[M>1, \theta<-3\Lambda(M)],$$
$$D_4 = \mathrm{E}[M>1, -3\Lambda(M)<\theta<\Lambda(M)].$$

(See fig. 1, p. 859.) The solutions defined in D_1 and D_3 were derived in the present paper, while those defined in D_2 and D_4 were derived in [**6**, § 11].

In the simplified case, using the theory of hypergeometric equations, it was possible to combine these representations into one, which yields solutions of (2.2) defined in the whole M-θ-plane. In the exact compressibility equation the problem remains of combining these four representations into one. This problem can be attacked by using the integral operator of the first kind in addition to that of the second kind, and, in analogy to the simplified case, developing a theory of differential equations with singular coefficients, which would furnish us with information corresponding to that used in the simplified case. As will be shown elsewhere, the methods of the Fuchs theory for ordinary differential equations can be generalized to the case of partial differential

equations of type (2.14) with F being an analytic function of two complex variables and possessing certain singularities. In particular, if the singularity surfaces are linear, the following results which will be proved in a subsequent paper, will be valid:

THEOREM 5.3. *Let the coefficient F of ψ in equation*

$$(5.35) \quad (\partial^2\psi/\partial Z_1{}^2) + (\partial^2\psi/\partial Z_2{}^2) + F\psi = 0, \quad Z_1 = \lambda + i\Lambda, \quad Z_2 = \theta + i\theta,$$

have the form

$$(5.36) \quad F = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} A_{m-1,n-1} W^{m-1} Z^{n-1}, \; A_{-1,-1} \neq 0, \qquad W = Z_1 + iZ_2, \quad Z = Z_1 - iZ_2$$

where the coefficients $A_{m,n}$ satisfy the inequalities

$$(5.37) \qquad | A_{m-1,n-1} | < A/\rho_1{}^m \rho_2{}^n, \quad A \geq 1, \quad \rho_1 > 0, \quad \rho_2 > 0,$$

being suitably chosen constants.

If $r \neq 0$ is a complex constant which satisfies the inequalities

$$(5.38a) \qquad r \neq -(m/2)[1 \pm (1 + A_{-1,-1}/mn)^{\frac{1}{2}}], \qquad 0 \leq \arg (1 + A_{-1,-1}/mn)^{\frac{1}{2}} < \pi/2,$$

$$(5.38b) \qquad r \neq A_{-1,-1}/4n,$$

$$(5.38c) \qquad r \neq -m$$

for $m = 0, 1, \cdots$; $n = 0, 1, \cdots$; $(m, n) \neq (0, 0)$, then the expression

$$(5.39) \qquad \psi = W^r Z^{-A_{-1,-1}/4r} \sum_{n=0}^{\infty} B_{mn} W^m Z^n,$$

will represent a solution of (5.35). Here B_{00} is an arbitrary complex constant, and the B_{mn} have to be determined from the equations

$$(5.40) \quad 4[mn + rn - (A_{-1,-1}m/4r)]B_{mn} = -\sum_{\mu=0}^{m-1} \sum_{\nu=0}^{n} A_{m-\mu-1,n-\nu-1} B_{\mu\nu} - \sum_{\nu=0}^{n-1} A_{-1,n-\nu-1} B_{m\nu}.$$

(5.38) *converges in*

$$\mathrm{E}[| W | < \rho_1/AB, \; | Z | < \rho_2/AB] \text{ where } B = \max [1, | B_{00} |, \mathrm{M}/4],$$

and

$$\mathrm{M} = \max \{(mn + m + n)/| mn + rn - 4^{-1}r^{-1}mA_{-1,-1} |\}, \qquad m = 0, 1, 2, \cdots, \; n = 0, 1, 2, \cdots, \; (mn) \neq (0, 0).$$

THEOREM 5.4. *Let the coefficient F of ψ in* (5.35) *have the form*

$$F = Z_1^{-2}A_{-2,0} + Z_1^{-1}A_{-1}(Z_2) + \sum_{n=0}^{\infty} A_n(Z_2)Z_1^n \tag{5.41}$$

where

$$A_{-2,0} \neq (1-n^2)/4,, \qquad n = 0, 1, 2, \cdots \tag{5.42}$$

and the $A_n(Z_2)$ are analytic functions of Z_2 which are dominated by $A/(R-Z_2)^{n+2}$, (i.e., for which $A_n(Z_2) \ll A/(R-Z_2)^{n+2}$ $n = -1, 0, 1, 2, \cdots$ holds) *A and R being suitably chosen positive constants. Then the expression*

$$\psi = Z_1^{r_s}\sum_{n=0}^{\infty} B_n(Z_2)Z_1^n, \qquad s = 1, 2, \tag{5.43}$$

represents a solution of (5.35) *which is defined in*

$$\mathrm{E}[\,|Z_1| < (R - |Z_2|)/(1+e), \quad |Z_2| < 1 - ((-1)^s(1-4A_{-2,0})^{\frac{1}{2}} + A)/(1-A_{-2,0})^{\frac{1}{2}}] \tag{5.44}$$

where

$$e = \max\,[\,|(1-(-1)^s(1-4A_{-2,0})^{\frac{1}{2}} + A)/(1-4A_{-2,0})^{\frac{1}{2}}\,|,\ |A/(1+(-1)^s(1-4A_{-2,0})^{\frac{1}{2}})|\,] \tag{5.45}$$

$$r_s = \tfrac{1}{2} + (-1)^s(\tfrac{1}{4} - A_{-2,0})^{\frac{1}{2}},\ 0 \leq \arg(\tfrac{1}{4} - A_{-2,0})^{\frac{1}{2}} < \pi/2, \qquad s = 1, 2.$$

Here $B_0(Z_2)$ is an arbitrary function for which

$$B_0(Z_2) \ll B(1-R^{-1}Z_2)^{-2}, \qquad B > 0, \tag{5.46}$$

and $B_n(Z_2)$ are to be determined by the relations

$$\begin{aligned}&[1+(-1)^s(1-4A_{-2,0})^{\frac{1}{2}}]B_1(Z_2) = -A_{-1}(Z_2)B_0(Z_2)\\ &n[n+(-1)^s(1-4A_{-2,0})^{\frac{1}{2}}]B_n(Z_2)\\ &\qquad = -[B''_{n-2}(Z_2) + \sum_{\nu=1}^{n} A_{\nu-2}(Z_2)B_{n-\nu}(Z_2)], \qquad n = 2, 3, \cdots.\end{aligned} \tag{5.47}$$

We should like to add here a remark regarding the general question of analytic continuation of a solution $\psi(Z, \bar{Z})$ of a linear partial differential equation. If ψ is given in two different domains, say B_1 and B_2 by different representations, say, in B_1 by the integral operator of the first kind in the form

$$\psi = \psi_1 \equiv \boldsymbol{p}[g(Z)] + \bar{\boldsymbol{p}}[h(\bar{Z})] \tag{5.48}$$

[see (3.11)] and in B_2 by another operator

$$\psi = \psi_2 \equiv \int_{t=-1}^{1} E(Z, \bar{Z}, t)f(\tfrac{1}{2}Z(1-t^2))\,(dt/(1-t^2)^{\frac{1}{2}}) \tag{5.49}$$

[see (3. 4a)] (not necessarily of the first kind), and the origin lies in $B_1 \cap B_2$, then the problem of the analytic continuation of ψ_2 into the domain B_1 is equivalent to the determination of g and h from a given $f = \sum_{n=0}^{\infty} \alpha_n Z^n$. Setting $Z = 0$, and $\bar{Z} = 0$, respectively, in the relation

$$(5.50) \qquad \boldsymbol{p}[g(Z)] + \bar{\boldsymbol{p}}[h(\bar{Z})] = \int_{t=-1}^{1} E(Z, \bar{Z}, t) f(\tfrac{1}{2}Z(1-t^2))\, dt/(1-t^2)^{\frac{1}{2}}$$

we obtain the identities

$$(5.51) \qquad \sum_{n=0}^{\infty} Z^n \sum_{\nu=0}^{n} \tau_{n\nu}{}^{(1)} \alpha_\nu = g(Z) + \bar{R}(0, Z) h(0),$$
$$\alpha_{00} \sum_{n=0}^{\infty} \bar{Z}^n \tau_{nn}{}^{(2)} = R(0, \bar{Z}) g(0) + h(\bar{Z})$$

where

$$(5.52) \qquad \tau_{n\nu}{}^{(1)} = \int_{t=-1}^{1} E_{n-\nu}{}^{(1)} (1-t^2)^{\nu-\frac{1}{2}} 2^{-\nu} dt,$$
$$\tau_{nn}{}^{(2)} = \int_{t=-1}^{1} E_n{}^{(2)}(t) (1-t^2)^{-\frac{1}{2}} dt,$$
$$E(Z, 0, t) = \sum_{n=0}^{\infty} E_n{}^{(1)}(t) Z^n, \qquad E(0, \bar{Z}, t) = \sum_{n=0}^{\infty} E_n{}^{(2)}(t) \bar{Z}^n.$$

[See **5**, page 310.] It is thus seen that the analytic continuation of ψ_2 into B_1 and all the problems arising from it, such as the determination of the singularities, etc., are reduced to similar problems *in the theory of functions of one complex variable* which are given by their power series expansions.

6. The determination of the associate function in terms of the given streamfunction. We now turn to the problem of determining the associate function f in terms of the given streamfunction. In the case of the integral operator of the first kind, the formulas (3. 7), (3. 10), (3. 15), (3. 16) yield the associate $g_1(Z)$. Therefore, if the continuation of the streamfunction (which a priori is given in the real plane) to complex values of the arguments is known, then these formulas immediately yield the associate. In the case of general integral operators (in particular, of that of the second kind), if the streamfunction is given in the form of a power series, then it may be shown (see [**5**], p. 310) that the associate can be expressed in the form of a power series. On the other hand, the streamfunction ψ is, in many instances, given in some different form, say the values of ψ and $\partial\psi/\partial M$ are given on a line $M = \text{const.}$ If these quantities are analytic functions of θ, then from these data it will be possible to determine the coefficients of the series development from which we may then determine the associate functions in the way indicated above. However, this procedure gives only the function

element of $f(Z)$, so that the function is in general determined only in a sufficiently small neighborhood of the interval of $M = \text{const.}$ along which the values $\psi = \chi_1(\theta)$ and $\partial\psi/\partial M = \chi_2(\theta)$ are given. We shall show in the present section that, in the case of certain generating functions of the second kind, a formula expressing f in terms of the values $\psi = \chi_1(\theta)$ and $\partial\psi/\partial M = \chi_2(\theta)$ on the sonic line, $M = 1$, can be derived. According to the consideration of **5** an integral operator of the second kind can be written in the form [17]

$$(6.1) \qquad \psi(\lambda,\theta) = \operatorname{Im}\Big[\int_{C_2} E(Z,\bar{Z},t) f(\tfrac{1}{2}Z(1-t^2))(1-t^2)^{-\frac{1}{2}}dt\Big]$$
$$= (2i)^{-1}\int_{C_2}[Ef - \bar{E}\bar{f}](1-t^2)^{-\frac{1}{2}}dt$$

where

$$(6.2) \qquad E = A_1 E^{(1)} + [\tfrac{1}{2}Z(1-t^2)]^{2/3} A_2 E^{(2)}, \qquad A_\kappa \text{ complex const.}$$

Here $E^{(\kappa)}$, $\kappa = 1, 2$, are the generating functions introduced in (5.5) and (4.1); C_2 is a simple curve in the complex t-plane which connects $t = -1$ with $t = 1$ and, except for the endpoints, lies outside $\mathrm{E}[\,|\,t\,| \leq 1]$. Moreover, C_2 has to be chosen in such a manner that $[\frac{1}{2}Z(1-t^2)]$ lies in the regularity domain of f for the values of Z under consideration.

We assume the associate function to be of the form

$$(6.3) \qquad f(\zeta) = \sum_{\nu=0}^{\infty} c_\nu \zeta^{\nu+1/6}, \qquad c_\nu \text{ complex const.}$$

which is suggested by previous considerations. Under the assumption that the (complex) constants A_1, A_2 satisfy the inequality

$$(6.4) \qquad \operatorname{Im}[A_2\bar{A}_1] \neq 0,$$

the desired inverse formula for f in terms of $\chi_1(\theta)$ and $\chi_2(\theta)$ is given in

THEOREM 6.1. *Let $\psi(\lambda,\theta)$ be a (real) solution of the compressibility equation* (2.8a) *which is defined in a domain, say B, situated in* $[3^{\frac{1}{2}}\,|\,\lambda\,| < |\,\theta\,|, \lambda < 0]$ *and such that its boundary includes an interval* $[\theta_0 \leq \theta \leq \theta_1]$ *of the transonic line $\lambda = 0$ not containing the origin. Let*

$$(6.5\text{a}) \qquad \lim_{\lambda\to 0^-}\psi(\lambda,\theta) = \chi_1(\theta) = \sum_{\nu=0}^{\infty} a_\nu^{(1)}\theta^\nu, \qquad a_\nu^{(\kappa)} \text{ real const.}$$

$$(6.5\text{b}) \qquad \lim_{\lambda\to 0^-}\psi_M(\lambda,\theta) = \chi_2(\theta) = 3^{1/3}(1-h^2)^{2/3}\sum_{\nu=0}^{\infty} a_\nu^{(2)}\theta^\nu, \; \psi_M = \partial\psi/\partial M,$$

(see (2.6)) *then the* P_2 *associate f of the integral operator* (6.1) *with generating function* (6.2) *and A_1, A_2 satisfying* (6.4) *is given by*

[17] In the remainder of this section we shall omit the subscript "2" in E_2, and shall write simply E.

$$(6.6)\quad f(\zeta) = -\{(-2i\zeta)^{1/6}/3^{1/2}\pi S_0{}^2\operatorname{Im}[A_2\bar{A}_1]\}[-\bar{d}_0\int_C t^{-1/3}\chi_1(\sigma)\,dt + \sum_{k=1}^{2}(-1)^k\bar{d}_k\int_C t^{-5/3}\chi_k(\sigma)\,dt]$$

where

$$(6.7)\qquad \sigma = -2i\zeta(1-t^2),$$

the constants d_ν, $\nu = 0, 1, 2$, *are*

$$(6.8)\quad d_0 = -(2/3)i^{3/2}S_0A_2,\quad d_1 = -(2^{5/3}/3)i^{1/6}S_0S_1A_1,\quad d_2 = -i^{1/6}S_0A_1,$$

the S_ν *are defined in* (4.4).

REMARK 6.1. Developing the right-hand side of (2.6) into an infinite series and inverting it, we obtain the following series for $(\partial\lambda/\partial M)$

$$(6.9)\quad \partial\lambda/\partial M = [3^{1/3}(1-h^2)^{2/3}](-\lambda)^{1/3} \times [1 + 3^{1/3}(1-h^2)^{-4/3}((3/10) + \tfrac{1}{2}h^2 - (4/5)h^4)(-\lambda)^{2/3} + \cdots]$$

Therefore it is legitimate to consider $\lim_{\lambda\to 0^-}(-\lambda)^{1/3}(\partial\psi/\partial\lambda)$, instead of $\lim_{\lambda\to 0^-}(\partial\psi/\partial M)$.

Proof. A formal computation yields (see (5.3), (5.5), (5.14), (6.8))

$$(6.10)\qquad \lim_{\lambda\to 0^-} E(Z,\bar{Z},t) = -d_2t^{-1/3}\theta^{-1/6}$$

and

$$(6.11)\quad \lim_{\lambda\to 0^-}(-\lambda)^{1/3}E_\lambda(Z,\bar{Z},t) = [d_1t^{-1/3} + d_0t^{-5/3}(1-t^2)^{2/3}]\theta^{-1/6}.$$

We now determine the limit values as $\lambda \to 0^-$ for the right-hand side of (6.1) and the derivative of this expression multiplied by $(-\lambda)^{1/3}$. Since by assumption these limit values are equal to expressions $\sum_{\nu=0}^{\infty} a_\nu{}^{(1)}\theta^\nu$ and $\sum_{\nu=0}^{\infty} 3^{1/3}(1-h^2)^{2/3}a_\nu{}^{(2)}\theta^\nu$ respectively, we obtain (by equating the coefficients of the θ^ν) the following system of equations

$$(6.12a)\qquad a_\nu{}^{(1)} = \operatorname{Im}[(-2i)^{-(\nu+1/6)}d_2I_\nu{}^{(1)}c_\nu],$$

$$(6.12b)\quad a_\nu{}^{(2)} = \operatorname{Im}[(-2i)^{-(\nu+1/6)}(d_1I_\nu{}^{(1)} + d_0I_\nu{}^{(2)})c_\nu],\quad \nu = 1, 2, \cdots$$

where

$$(6.13a)\quad I_\nu{}^{(1)} = \int_{C_2} t^{-1/3}(1-t^2)^{\nu-1/3}dt = \pm\tfrac{1}{2}e^{-(4/3)n\pi i}(e^{-(4/3)\pi i}-1)\frac{\Gamma(1/3)\Gamma(\nu+2/3)}{\Gamma(\nu+1)},\quad n = 0, \pm 1$$

$$(6.13b)\quad I_\nu{}^{(2)} = \int_{C_2} t^{-5/3}(1-t^2)^{\nu+1/3}dt = \pm\tfrac{1}{2}e^{-(2/3)n\pi i}(e^{-(2/3)\pi i}-1)\frac{\Gamma(-1/3)\Gamma(\nu+4/3)}{\Gamma(\nu+1)}.$$

We note that the last term on the right hand side of (6.13b) is obtained as follows:

$$I_\nu^{(2)} = -(3/2)\int_{C_2} (1-t^2)^{\nu+1/3} d(t^{-2/3})$$

$$= -(3/2)[t^{-2/3}(1-t^2)^{\nu+1/3}]_{t=-1}^{t=-1} - 3(\nu+1/3)\int_{C_2} t^{1/3}(1-t^2)^{\nu-2/3}dt.$$

The first term in the last expression vanishes and in the second term, as in $I_\nu^{(1)}$, we replace the integration curve C_2 by the segments $(-1, 0^-)$ and $(0^+, 1)$ and a half-circle around the origin.

REMARK. Since the curve C_2 need only satisfy the inequality $|t| > |2\lambda/(\lambda + i\theta)|^{\frac{1}{2}}$ (see 4.11), it is valid to replace it by a curve consisting of the segments $(-1, 0^-)$ and $(0^+, 1)$ and a half-circle around the origin provided the radius of the latter is greater than $|2\lambda/(\lambda + i\theta)|^{\frac{1}{2}}$ which approaches zero as $\lambda \to 0$. The right hand sides of (6.13a) and (6.13b) are obtained by making the substitution $t^2 = \tau$ and considering the integrals in the three-sheeted τ-plane. In the following, we choose the sheet for which $n = 1$.

We introduce the integrals

$$\text{(6.14a)} \quad J_\nu^{(1)} = \int_{C_2} t^{-5/3}(1-t^2)^\nu dt$$

$$= \pm \tfrac{1}{2} e^{-(2/3)n\pi i}(e^{-(2/3)\pi i} - 1)\frac{\Gamma(-1/3)\Gamma(\nu+1)}{\Gamma(\nu+2/3)}$$

$$\text{(6.14b)} \quad J_\nu^{(2)} = \int_{C_2} t^{-1/3}(1-t^2)^\nu dt$$

$$= \pm \tfrac{1}{2} e^{-(4/3)n\pi i}(e^{-(4/3)\pi i} - 1)\frac{\Gamma(1/3)\Gamma(\nu+1)}{\Gamma(\nu+4/3)}$$

and note that

$$\text{(6.15)} \qquad I_\nu^{(1)} J_\nu^{(1)} = I_\nu^{(2)} J_\nu^{(2)} = -3^{3/2}2^{-1}\pi.$$

The determinant of the system (6.12) does not vanish for any ν since it equals

$$\text{(6.16)} \qquad -(i/2)\,\mathrm{Im}\,[d_0\bar{d}_2 I^{(2)}\overline{I^{(1)}}]$$

which would vanish only if $\mathrm{Im}[d_0\bar{d}_2] = 0$. Since, by (6.8), $\mathrm{Im}[d_0\bar{d}_2] = (2/3)S_0^2\,\mathrm{Im}[A_2\bar{A}_1]$, the determinant does not vanish, because of condition (6.4).

Solving (6.2), we obtain

$$(6.17)\quad c_\nu = \frac{-(-2i)^{\nu+1/6}}{(3^{\frac{1}{3}}\pi S_0^2)\,\mathrm{Im}[A_2\bar{A}_1]}\left[-a_\nu^{(1)}\bar{d}_0 J_\nu^{(1)} + \sum_{k=1}^{2}(-1)^k a_\nu^{(k)}\bar{d}_k J_\nu^{(2)}\right]$$

from which (6.6) follows.

HARVARD UNIVERSITY.

BIBLIOGRAPHY.

1. Stefan Bergman, "Zur Theorie der Funktionen, die eine lineare partielle Differentialgleichung befriedigen, I," *Recueil Mathématique*, new series, vol. 2 (1937), pp. 1169-1198.
2. ———, "The approximation of functions satisfying a linear partial differential equation," *Duke Mathematical Journal*, vol. 6 (1940), pp. 537-561.
3. ———, "Linear operators in the theory of partial differential equations," *Transactions of the American Mathematical Society*, vol. 53 (1943), pp. 130-155.
4. ———, "Solutions of linear partial differential equations of the fourth order," *Duke Mathematical Journal*, vol. 11 (1944), pp. 617-649.
5. ———, "Certain classes of analytic functions of two real variables and their properties," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 299-331.
6. ———, "On two-dimensional flows of compressible fluids," *National Advisory Committee for Aeronautics, Technical Note No. 972*, 1945.
7. ———, "Methods for the determination and computation of flow patterns of a compressible fluid," *National Advisory Committee for Aeronautics, Technical Note No. 1018*, 1946.
8. ———, "On supersonic and partially supersonic flows, *National Advisory Committee for Aeronautics, Technical Note No. 1096*, 1946.

8a. ———, "Two-dimensional subsonic flows of a compressible fluid and their singularities," *Transactions of the American Mathematical Society*, vol. 61 (1947), pp. 452-498.

9. ———, "On tables for the determination of transonic flow patterns," *Hans Reissner Anniversary Volume*, 1949.
10. ———, "On functions satisfying certain partial differential equations of elliptic type and their representation," *Duke Mathematical Journal*, vol. 14 (1947), pp. 349-366.
11. S. A. Chaplygin, "On gas jets," *Scientific Memoirs, Moscow University, Mathematics-Physics Section*, vol. 21 (1902), pp. 1-121. (Eng. trans., pub. by NACA TM No. 1063, 1944 and also pub. by Brown Univ., 1944.)
12. A. R. Forsyth, *A treatise on differential equations*, London (1943).
13. F. Frankl, "On the problems of Chaplygin for mixed sub- and supersonic flows," *Bulletin de L'Académie des Sciences de L'URSS*, vol. 9 (1945), pp. 121-143.
14. E. W. Hobson, *The theory of spherical and ellipsoidal harmonics*, Cambridge (1931).
15. Thomas M. MacRobert, *Functions of a complex variable*, London (1933).
16. F. Tricomi, "Sulle equazioni lineari alle derivato parziali di 2° ordine di tipo misto," *Memorie della Reale Accademia Nazionale dei Lincei*, vol. 14 (1923), pp. 133-247.

ASYMPTOTIC RELATIONS FOR THE EIGENFUNCTIONS OF CERTAIN BOUNDARY PROBLEMS OF POLAR TYPE.*

By Åke Pleijel.

Introduction. Let V be a bounded connected and open domain in the Euclidean (x_1, x_2, x_3)-space and let S be its boundary. We consider the equation

$$\Delta u - p(x_1, x_2, x_3) u + \lambda q(x_1, x_2, x_3) u = 0 \text{ in } V \tag{1}$$

with one of the boundary conditions

$$u = 0 \text{ on } S \tag{2}$$

or

$$\partial u / \partial n = 0 \text{ on } S. \tag{3}$$

In (1)

$$\Delta = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2 + \partial^2/\partial x_3^2$$

is the Laplace operator, $p(x_1, x_2, x_3)$ and $q(x_1, x_2, x_3)$ are given real functions and λ is a parameter. In (3), $\partial u/\partial n$ is the derivative of u in the direction of the normal of S. The boundary conditions (2) and (3) shall be taken in R. Courant's generalized sense and when (3) is concerned we assume $V + S$ to be of the type considered by this author in his investigations on eigenvalue-problems of the kind (1), (3); see [3].[1] *In the main part of this paper we assume that the function p has a positive lower bound in V.* In 12 we indicate how our results can be shown to subsist even under the more general condition that p is non-negative. Although less restrictive regularity conditions would suffice for our discussion *we assume the functions p and q to be bounded and continuous with their first-order derivatives.*

As is well known the eigenvalues of each of our problems (values of λ for which the problem admits of a solution $\not\equiv 0$) are all real and have no finite limit-points. If q is non-negative (and $\not\equiv 0$) there are an infinity of positive but no negative eigenvalues. *The problems are said to be of polar type if the function q changes its sign in V.* In this case there are an infinity of positive eigenvalues and an infinity of negative ones. We denote these eigenvalues by λ_n, $n = \cdots -2, -1, 0, 1, 2, \cdots$, and the corre-

* Received March 25, 1948.

[1] Number in brackets refer to the bibliography at the end of the paper.

sponding eigenfunctions by $\phi_n(x_1, x_2, x_3)$. The sequence of eigenvalues is supposed to be orderd so that $\lambda_i \leq \lambda_k$ if $i < k$, $\lambda_i < 0$ if $i < 0$ and $\lambda_i \geq 0$ if $i \geq 0$. The eigenfunctions shall be orthonormalized so as to fulfill the conditions

$$\int_V q(P)\phi_i(P)\phi_k(P)\,dV = \begin{cases} \operatorname{sign}\lambda_i & \text{if } i = k \\ 0 & \text{if } i \neq k \end{cases}.$$

where $P = (x_1, x_2, x_3)$ and $dV = dx_1 dx_2 dx_3$. We introduce the functions

$$q_+(P) = \begin{cases} q(P) & \text{if } q(P) > 0 \\ 0 & \text{if } q(P) \leq 0 \end{cases}$$

and

$$q_-(P) = \begin{cases} -q(P) & \text{if } q(P) < 0 \\ 0 & \text{if } q(P) \geq 0. \end{cases}$$

If $N(0 < \lambda_n < T)$ and $N(-T < \lambda_n < 0)$ are the numbers of eigenvalues in the assigned intervals the asymptotic relations [2]

$$\text{(4)} \qquad N(0 < \lambda_n < T) = [\int_V q_+^{3/2}(P)\,dV/6\pi^2]T^{3/2} + o(T^{3/2})$$

and

$$\text{(5)} \qquad N(-T < \lambda_n < 0) = [\int_V q_-^{3/2}(P)\,dV/6\pi^2]T^{3/2} + o(T^{3/2})$$

are known to hold true for each of the considered problems when T tends to $+\infty$ (see [5]).

Following a method due to T. Carleman ([1], [2]) *we deduce the asymptotic relations for the eigenfunctions*

$$\text{(6)} \qquad \sum_{0<\lambda_n<T} \phi_n^2(P) = [q_+^{\frac{1}{2}}(P)/6\pi^2]T^{3/2} + o(T^{3/2})$$

$$\text{(7)} \qquad \sum_{-T<\lambda_n<0} \phi_n^2(P) = [q_-^{\frac{1}{2}}(P)/6\pi^2]T^{3/2} + o(T^{3/2})$$

valid when $q(P) \neq 0$. Let us remark that after a multiplication by $q(P)$ a formal but evidently non-legitimate integration of (6) and (7) gives the asymptotic formulas for the eigenvalues.

Let P and Q be points in V and let z be a complex number. We denote by $G(P, Q; z)$ the Green's function which for $P = Q$ is singular as $1/4\pi r_{PQ}$ and which as a function of one of its argument points satisfies the equation $\Delta u - pu + zqu = 0$ in V and the boundary condition (2) or (3) on S.

---

[2] $f(T) = o(g(T))$ when $T \to +\infty$ means that $\lim_{T=+\infty} f(T)/g(T) = 0$.

$G(P,Q;z)$ exists if z is not equal to an eigenvalue. We put $G(P,Q;0) = G(P,Q)$. *If q has a positive lower bound in V* we have

$$(8) \qquad G(P,Q;z) - G(P,Q) = z \sum_{n=0}^{\infty} \phi_n(P)\phi_n(Q)/\lambda_n(\lambda_n - z)$$

where λ_n, $n = 0, 1, 2, \cdots$, are the positive eigenvalues of the considered problem. (If $p \equiv 0$ and if the boundary condition (3) is considered, the value $z = 0$ is an eigenvalue; in this case, (8) should be replaced by a similar development of $G(P,Q;z) - G(P,Q;z_0)$ where z_0 is no eigenvalue). If for z real and equal to $-r$ the asymptotic behaviour of

$$(9) \qquad \lim_{Q=P} [G(P,Q;z) - G(P,Q)]$$

is deduced for r tending to $+\infty$ one obtains from (8) an asymptotic formula for the sum

$$\sum_{n=0}^{\infty} \phi_n{}^2(P)/\lambda_n(\lambda_n + r).$$

By the use of a Tauberian theorem (see **9**) it is then possible to prove the relation

$$\sum_{\lambda_n < T} \phi_n{}^2(P) = [q^{\frac{3}{2}}(P)/6\pi^2]T^{3/2} + o(T^{3/2})$$

valid when the function q has a positive lower bound in V.

In attempting to apply this method *in the case when q changes its sign in V* we cannot let z tend to infinity through real values but have to consider the expression (9) for large complex values of z. Actually we shall study (9) when z tends to infinity through imaginary values. We do this by considering certain variational problems of saddle-point type. By the help of an asymptotic formula for (9) obtained in this way we verify that the development

$$(10) \qquad \lim_{Q=P} [G(P,Q;z) - G(P,Q)] = \sum_{n=-\infty}^{+\infty} \phi_n{}^2(P)/\lambda_n(\lambda_n - z)$$

is valid when $q(P) \neq 0$. Hence the asymptotic behaviour of the series in (10) is known when z tends to infinity through imaginary values and when, at the point P, $q(P) \neq 0$. By a suitable use of the Tauberian theorem the relations (6) and (7) follow.

1. The Green's function. For a fixed but arbitrary position of the point P in V we choose the singular part $\Gamma(P,Q;z)$ of the Green's function

$$G(P,Q;z) = \Gamma(P,Q;z) - \gamma(P,Q;z)$$

so as to fulfill the following conditions. *The difference*

$$u(Q) = \Gamma(P, Q; z) - 1/4\pi r_{PQ}$$

shall be continuous having continuous derivatives of the first and second order with respect to Q; *the integral*

$$\int_V (\mathrm{grad}^2 u + pu^2)\, dV$$

shall be finite. We consider a sphere $V' = (r_{PQ} < R)$ *with center in* P *which together with its boundary* $S' = (r_{PQ} = R)$ *shall be contained in* V. *Outside of* V' *the function* $\Gamma(P, Q; z)$ *shall vanish identically.* The explicit construction of $\Gamma(P, Q; z)$ will be performed in **6**.

As $\Gamma(P, Q; z)$ and its derivatives vanish when Q approaches S, the regular parts $\gamma(P, Q; z)$ of the Green's functions fulfill the conditions

$$(\Delta - p + zq)_Q \gamma(P, Q; z) = (\Delta - p + zq)_Q \Gamma(P, Q; z) \text{ in } V$$

and

$$\gamma(P, Q; z) = 0 \text{ when } Q \text{ lies in } S$$

or

$$\partial\gamma(P, Q; z)/\partial n = 0 \text{ when } Q \text{ lies in } S.$$

2. Introduction of certain integral forms. Under suitable regularity conditions on u and v which do not necessarly involve continuity at S' we obtain from

$$D(u, v) = \int_V (\mathrm{grad}\, u\, \mathrm{grad}\, v + puv - zquv)\, dV$$

by partial integration the expression

$$D^*(u, v) = -\int_V u(\Delta v - pv + zqv)\, dV$$
$$+ \int_S u(\partial v/\partial n)\, dS + \int_{S'_i} u(\partial v/\partial n)\, dS + \int_{S'_e} u(\partial v/\partial n)\, dS.$$

Here S'_i and S'_e both coincide with S'. But S'_i is considered as the boundary of V', its normal is directed outwards from V' and the values of the functions in the integral over S'_i are obtained by approaching S' from the interior of V'. In the same way S'_e is considered as one part of the boundary $S + S'$ of $V - V'$; the normals of S and S'_e are directed outwards from $V - V'$ and in the integrals over S and S'_e the values of the integrands are obtained by approaching S and S' from the interior of $V - V'$.

The relation

$$D^*(\Gamma(P,Q;z),u(Q)) - D^*(u(Q),\Gamma(P,Q;z)) = u(P)$$

is easily seen to be valid if u is sufficiently regular and in particular for $u(Q) = \gamma(P,Q;z)$. With these values of $u(Q)$ the identity

$$\begin{aligned} D^*(u,u) - 2D^*(u,\Gamma) + D^*(\Gamma,\Gamma) - D^*(\Gamma - u, \Gamma - u) \\ = D^*(\Gamma,u) - D^*(u,\Gamma) \end{aligned}$$

gives

$$\begin{aligned} D^*(\gamma,\gamma) - 2D^*(\gamma,\Gamma) + D^*(\Gamma,\Gamma) - D^*(\Gamma-\gamma,\Gamma-\gamma) \\ = \gamma(P,P;z). \end{aligned}$$

Here $D^*(\gamma,\gamma)$ equals $D(\gamma,\gamma)$ and $D^*(\Gamma-\gamma,\Gamma-\gamma)$ vanishes. Thus for the regular parts of our Green's functions we obtain the formula

$$E(\gamma;\Gamma) + D^*(\Gamma,\Gamma) = \gamma(P,P;z) \tag{11}$$

where $E(u;\Gamma) = D(u,u) - 2D^*(u,\Gamma)$.

By splitting real and imaginary parts we write $z = x + iy$, $u = u_1 + iu_2$, $G = G_1 + iG_2$, $\Gamma = \Gamma_1 + i\Gamma_2$, $\gamma = \gamma_1 + i\gamma_2$ and

$$\begin{aligned} D(u,v) &= D_1(u_1,u_2;v_1,v_2) + iD_2(u_1,u_2;v_1,v_2), \\ D^*(u,v) &= D^*_1(u_1,u_2;v_1,v_2) + iD^*_2(u_1,u_2;v_1,v_2), \\ E(u;\Gamma) &= E_1(u_1,u_2;\Gamma_1,\Gamma_2) + iE_2(u_1,u_2;\Gamma_1,\Gamma_2). \end{aligned}$$

The relation (11) is equivalent to the two formulas

$$E_1(\gamma,\gamma_2;\Gamma_1,\Gamma_2) + D^*_1(\Gamma_1,\Gamma_2;\Gamma_1,\Gamma_2) = \gamma_1(P,P;z), \tag{12}$$

$$E_2(\gamma,\gamma_2;\Gamma_1,\Gamma_2) + D^*_2(\Gamma_1,\Gamma_2;\Gamma_1,\Gamma_2) = \gamma_2(P,P;z). \tag{13}$$

With $zq(Q) = h_1(Q) + ih_2(Q)$ we have

$$\begin{aligned} E_1(u_1,u_2;\Gamma_1,\Gamma_2) = \int_V \{&\text{grad}^2 u_1 + pu_1^2 - \text{grad}^2 u_2 - pu_2^2 \\ &- h_1(u_1^2 - u_2^2) + 2h_2u_1u_2 \\ &+ 2u_1(\Delta\Gamma_1 - p\Gamma_1 + h_1\Gamma_1 - h_2\Gamma_2) \\ &- 2u_2(\Delta\Gamma_2 - p\Gamma_2 + h_1\Gamma_2 + h_2\Gamma_1)\}dV. \end{aligned} \tag{14}$$

3. Variational problems. We denote by $\mathfrak{D}'$ the class of functions u which have the following properties. They are continuous and admit piecewise-continuous first-order derivatives in $V - S'$. The integral

$$(u, u) = \int_V (\text{grad}^2 u + pu^2)\,dV$$

exists. In the case of the problem (1), (2) the functions of $\mathfrak{D}'$ vanish on S in Courant's generalized sense.

We denote by $(\mathfrak{D}')$ the sub-class of $\mathfrak{D}'$ containing functions which are continuous and have piecewise-continuous first-order derivatives in V.

Putting $x = 0$ (which gives $h_1(Q) \equiv 0$) we consider the following type of variational problems connected with the form $E_1(u_1, u_2; \Gamma_1, \Gamma_2)$.

(A) *With fixed values of the function u_2 and varying u_1 we search for* $M(u_2) = \min. E_1(u_1, u_2; \Gamma_1, \Gamma_2)$.

(B) *Varying u_2 we search for* $M = \max. M(u_2)$.

We distinguish five problems I, II, III, IV, V of this kind by the conditions on u_1 and u_2:

I	II	III	IV	IV
$u_1 = 0$ in V	$u_1 \,\varepsilon\, (\mathfrak{D}')$	$u_1 \,\varepsilon\, (\mathfrak{D}')$	$u_1 \,\varepsilon\, \mathfrak{D}'$	$u_1 \,\varepsilon\, \mathfrak{D}'$
$u_2 \,\varepsilon\, \mathfrak{D}'$	$u_2 \,\varepsilon\, \mathfrak{D}'$	$u_2 \,\varepsilon\, (\mathfrak{D}')$	$u_2 \,\varepsilon\, (\mathfrak{D}')$	$u_2 = 0$ in V

We call the values $M(u_2)$ obtained in these problems $M^I(u_2)$, $M^{II}(u_2)$, $\cdots M^V(u_2)$. By $\{M^I(u_2)\}$, $\{M^{II}(u_2)\}, \cdots \{M^V(u_2)\}$ we denote the sets of values $M^I(u_2), M^{II}(u_2), \cdots M^V(u_2)$. The least upper bounds in these sets are $M^I, M^{II}, \cdots M^V$.

Evidently $M^I(u_2) \geq M^{II}(u_2)$ for the conditions on u_1 are stronger in problem I than in problem II. It follows that $M^I \geqq M^{II}$. In problem III the class of admissible functions u_2 is part of the corresponding class in problem II which shows that $\{M^{II}(u_2)\} \supseteq \{M^{III}(u_2)\}$. This gives $M^{II} \geq M^{III}$. Continuing in this way we obtain

$$MI \geqq M^{II} \geqq M^{III} \geq M^{IV} \geq M^V$$

of which however we use only

$$M^I \geqq M^{III} \geqq M^V. \tag{15}$$

Note. It follows from $| q(P) | \leq [\max. | q |/\min. p] p(P)$ that

$$\left| \int_V h_2 u_1 u_2 dV \right| \leq \text{const. } (u_1, u_1)^{\frac{1}{2}} (u_2, u_2)^{\frac{1}{2}},$$

$$\left| \int_V u_1(\Delta\Gamma_1 - p\Gamma_1 - h_2\Gamma_2)\,dV \right| \leq \text{const. } (u_1, u_1)^{\frac{1}{2}}, \tag{16}$$

$$\left| \int_V u_2(\Delta\Gamma_2 - p\Gamma_2 + h_2\Gamma_1)\,dV \right| \leq \text{const. } (u_2, u_2)^{\frac{1}{2}}.$$

By the help of these inequalities it is easily seen that $M(u_2)$ in the problem (**A**), (**B**) is $> -\infty$. For an arbitrary but fixed u^*_1, $M(u_2) \leqq E_1(u^*_1, u_2; \Gamma_1, \Gamma_2)$. From (16) follows that $\max E_1(u^*_1; u_2; \Gamma_1, \Gamma_2)$ has a finite upper bound when this maximum is calculated with u^*_1 fixed. Thus $M = \max M(u_2)$ obtained in the problem (**A**), (**B**) is finite.

4. The values of M^I and M^V. In problem I the class u_1 consists only of the function $u_1 \equiv 0$. Hence $E_1(u_1, u_2; \Gamma_1, \Gamma_2)$ reduces to

$$(17)\quad E_1(0, u_2; \Gamma_1, \Gamma_2) = -\int_V [\operatorname{grad}^2 u_2 + pu_2^2 + 2u_2(\Delta\Gamma_2 - p\Gamma_2 + h_2\Gamma_1)]\, dV$$

and in order to solve our problem I we have to search for the maximum of this expression when $u_2 \varepsilon \mathfrak{D}'$, which leads to the equations

$$(18)\quad \left.\begin{aligned} &\Delta u_2 - pu_2 = \Delta\Gamma_2 - p\Gamma_2 + h_2\Gamma_1 \text{ in } V' \\ &\partial u_2/\partial n = 0 \text{ on } S', \end{aligned}\right\}$$

and

$$(19)\quad \left.\begin{aligned} &\Delta u_2 - pu_2 = 0 \text{ in } V - V', \\ &\partial u_2/\partial n = 0 \text{ on } S', \\ &u_2 = 0 \text{ on } S \text{ or } \partial u_2/\partial n = 0 \text{ on } S. \end{aligned}\right\}$$

The last set of equations (19) give $u_2 = 0$ in $V - V'$. And if $G^*(Q, \Pi)$ is the Green's function which belongs to the equation $\Delta u - pu = 0$ in V' and to the boundary condition $\partial u_2/\partial n = 0$ on S' the solution of the first set (18) is

$$(20)\qquad u_2(Q) = -\int_{V'} G^*(Q, \Pi)\,(\Delta\Gamma_2 - p\Gamma_2 + h_2\Gamma_1)_{P_\Pi}\, dV_\Pi.$$

With this function $u_2(Q)$ $(u_2(Q) = 0$ in $V - V')$ we have

$$M^I = E_1(0, u_2; \Gamma_1, \Gamma_2).$$

On account of (18) we obtain by partial integration

$$(21)\qquad M^I = -\int_{V'} u_2(\Delta\Gamma_2 - p\Gamma_2 + h_2\Gamma_1)\, dV.$$

For M^V we find similarly

$$(22)\qquad M^V = \int_{V'} u_1(\Delta\Gamma_1 - p\Gamma_1 - h_2\Gamma_2)\, dV$$

where

$$(23)\qquad u_1(Q) = -\int_{V'} G^*(Q, \Pi)\,(\Delta\Gamma_1 - p\Gamma_1 - h_2\Gamma_2)_{P_\Pi}\, dV_\Pi.$$

5. The value of M^{III}. A simple way to handle problem III is to consider the corresponding problem in $(\mathfrak{D}) \times (\mathfrak{D})$ where $(\mathfrak{D})$ is the Hilbert space obtained by completing $(\mathfrak{D}')$ with respect to the metric

$$(u, u) = \int_V (\operatorname{grad}^2 u + pu^2)\, dV.$$

Taking into account the inequalities (16) we see that the expression (according to our choice $x = 0$ we have $h_1 \equiv 0$)

$$E_1(u_1, u_2; \Gamma_1, \Gamma_2) = \int_V [\operatorname{grad}^2 u_1 + pu_1^2 - \operatorname{grad}^2 u_2 - pu_2^2 + 2h_2 u_1 u_2 + 2u_1(\Delta\Gamma_1 - p\Gamma_1 - h_2\Gamma_2) - 2u_2(\Delta\Gamma_2 - p\Gamma_2 + h_2\Gamma_1)]\, dV$$

defines a non-homogeneous quadratic form in $(\mathfrak{D}) \times (\mathfrak{D})$ which can be written

$$E_1(u_1, u_2; \Gamma_1, \Gamma_2) = (u_1, u_1) - (u_2, u_2) + 2(\boldsymbol{H}u_1, u_2) + 2(u_1, f_1) - 2(u_2, f_2) \tag{24}$$

where $\boldsymbol{H}$ is a linear, self-adjoint and bounded operator in $(\mathfrak{D})$ and where f_1 and f_2 are elements in $(\mathfrak{D})$ uniquely determined by the functions $(\Delta\Gamma_1 - p\Gamma_1 - h_2\Gamma_2)$ and $(\Delta\Gamma_2 - p\Gamma_2 + h_2\Gamma_1)$. The problem (**A**), (**B**) for the form (24) has a unique solution in $(\mathfrak{D}) \times (\mathfrak{D})$ which at the same time is the unique solution of the problem to find a $u_1 \times u_2$ in $(\mathfrak{D}) \times (\mathfrak{D})$ for which the variation of (24) vanishes. This problem, however, is solved by $u_1 = \gamma_1$, $u_2 = \gamma_2$ where $\gamma_1 + i\gamma_2$ is the regular part of the Green's function $G(P, Q; iy)$. γ_1 and γ_2 can evidently be considered as elements of $(\mathfrak{D})$. Hence,

$$M^{III} = E_1(\gamma_1, \gamma_2; \Gamma_1, \Gamma_2)$$

or according to (12)

$$M^{III} = \gamma_1(P, P; iy) - D^*{}_1(\Gamma_1, \Gamma_2; \Gamma_1, \Gamma_2).$$

From this formula and from the inequalities (15) it follows that

$$M^I + D^*{}_1(\Gamma_1, \Gamma_2; \Gamma_1, \Gamma_2) \geq \gamma_1(P, P; iy) \geq M^V + D^*{}_1(\Gamma_1, \Gamma_2; \Gamma_1, \Gamma_2). \tag{26}$$

6. Estimation of $\gamma_1(P, P; iy)$. We choose

$$\Gamma(P, Q; z) = [\eta_R(r_{PQ})/4\pi r_{PQ}] \exp\{-(p(P) - zq(P))^{\frac{1}{2}} r_{PQ}\} \tag{27}$$

where

$$\eta_R(r) = \begin{cases} 1 \text{ when } r \leq \frac{1}{2}R, \\ 1 - X(r)/X(R) \text{ when } \frac{1}{2}R < r \leq R, \\ 0 \text{ when } R < r, \end{cases}$$

and

$$X(r) = \int_{\frac{1}{2}R}^{r} (t - \tfrac{1}{2}R)^2 (t - R)^2 dt.$$

Evidently

$$(28) \qquad |\eta_R(r)| \leq 1, \quad |d\eta_R(r)/dr| \leq \text{const. } R^{-1}, \quad |d^2\eta_R(r)/dr^2| \leq \text{const. } R^{-2}.$$

The radius R of the sphere around P, $V' = (r_{PQ} < R)$, shall be so small that the sphere together with its boundary lies entirely in V. The square-root in (27) shall have a positive real part. We put $z = iy$.

Provided $q(P) \neq 0$ *at the fixed point* P [3]

$$(29) \qquad \Gamma(P, Q; iy) = O(r_{PQ}^{-1} \exp(-cy^{\frac{1}{2}} r_{PQ})) \text{ when } y \to +\infty,\ r_{PQ} \leq R,$$

where c is a positive constant (independent of Q and y). When $r \leq \frac{1}{2}R$ the function η_R is constant $= 1$ and

$$(\Delta_Q - p(Q) + iyq(Q))\Gamma(P, Q; iy) = (p(P) - p(Q))\Gamma - iy(q(P) - q(Q))\Gamma.$$

Hence,

$$(30) \qquad (\Delta_Q - p(Q) + iyq(Q))\Gamma = O(y \exp(-cy^{\frac{1}{2}} r_{PQ}))$$

which by the help of the inequalities (28) is easily seen to hold true throughout the interval $0 \leq r \leq R$.

On account of (29) and (30) we find that

$$(31) \qquad D^*(\Gamma, \Gamma) = O(y \int_0^R \exp(-2cy^{\frac{1}{2}} r) r\, dr) = O(1).$$

The Green's function $G^*(Q, \Pi)$ fulfills the relation

$$G^*(Q, \Pi) = O(r_{Q\Pi}^{-1}), \qquad Q \text{ and } \Pi \text{ in } V'.$$

On account of this and the evident relation

$$y \exp(-cy^{\frac{1}{2}} r) = O(y^{-d} r^{-2-2d})$$

where d is an arbitrarily small positive quantity we find for the function (20)

$$u_2(Q) = O(y^{-d} \int_{V'} r_{Q\Pi}^{-1} r_{P\Pi}^{-2-2d}\, dV_\Pi).$$

[3] As usual $f = O(g)$ means that f/g is bounded when the variables range in prescribed ways or in indicated domains.

holds true, the theorem asserts that

$$(42)\qquad \sum_{\mu_n < u} B_n = [H \sin(\pi b)/\pi(1-b)]u^{1-b} + o(u^{1-b})$$

for u tending to infinity.

Putting $\lambda_n^2 = \mu_n$ and $y^2 = r$ in (40) we get a formula of the type (41) and the Tauberian theorem gives

$$\sum_{|\lambda_n| < T} \phi_n^2(P)/|\lambda_n| = [\{|q(P)|\}^{\frac{3}{2}}/2\pi^2]T^{\frac{1}{2}} + o(T^{\frac{1}{2}})$$

when T tends to $+\infty$. It follows that the series

$$(43)\qquad \sum_{n=-\infty}^{\infty} \phi_n^2(P)/|\lambda_n|^{3/2+\epsilon}$$

is convergent for every positive ϵ. Thus $\sum_{n=-\infty}^{+\infty} \phi_n^2(P)/\lambda_n^2$ converges and (39) can be written in the form

$$(44)\qquad \begin{aligned} &\Big[\int_V q(\Pi)G^2(P,\Pi)\,dV - \sum_{n=-\infty}^{+\infty} \phi_n^2(P)/\lambda_n^2 \operatorname{sign}\lambda_n\Big] \\ &\quad + \sum_{n=-\infty}^{+\infty} \phi_n^2(P)/|\lambda_n|(\lambda_n - iy) \\ &\quad = [(\operatorname{sign} q(P) + i)\{\tfrac{1}{2}\ |q(P)|\}^{\frac{3}{2}}/4\pi]y^{-\frac{1}{2}} + O(y^{-1}). \end{aligned}$$

From the convergence of (43) it is easily seen that the last sum of the left-hand side of (44) tends to zero when y tends to $+\infty$. It follows that the expression in brackets on the left of (44) is zero and (44) reduces to

$$(45)\qquad \begin{aligned} &\sum_{n=-\infty}^{+\infty} \phi_n^2(P)/|\lambda_n|(\lambda_n - iy) \\ &\quad = [(\operatorname{sign} q(P) + i)\{\tfrac{1}{2}\,|q(P)|\}^{\frac{3}{2}}/4\pi]y^{-\frac{1}{2}} + O(y^{-1}). \end{aligned}$$

10. Consequences of the Tauberian theorem. From the Tauberian theorem of the preceding paragraph we deduce results to be used in connection with the formula (45).

Let (λ_n), $n = \cdots -2, -1, 0, 1, 2, \cdots$, be a sequence of real numbers such that $\lambda_i \leqq \lambda_k$ when $i < k$, $\lambda_i < 0$ when $i < 0$, $\lambda_i \geqq 0$ when $i \geqq 0$, $\lim_{n=-\infty} \lambda_n = -\infty$ and $\lim_{n=+\infty} \lambda_n = +\infty$. Let A_n be positive numbers such that $\sum_{n=-\infty}^{+\infty} A_n/|\lambda_n|$ converges. *We assume the relation*

$$(46)\qquad \sum_{n=-\infty}^{+\infty} A_n/(\lambda_n - iy) = (H_1 + iH_2)y^{-a} + o(y^{-a}),$$

$(a$ *real*, $0 < a < 1)$

to be valid when y *tends to* $+\infty$. By splitting real and imaginary parts we obtain from (46) when y tends to $+\infty$

$$(47)\qquad \sum_{n=-\infty}^{\infty} \lambda_n A_n/(\lambda_n^2 + y^2) = H_1 y^{-a} + o(y^{-a});$$

$$(48)\qquad \sum_{n=-\infty}^{\infty} A_n/(\lambda_n^2 + y^2) = H_2 y^{-a-1} + o(y^{-a-1}).$$

We put $\lambda_n^2 = \mu_n$ and $y^2 = r$ and get from (48) a formula of the type (41). By the help of the Tauberian theorem we deduce corresponding to (42) a formula describing the asymptotic behaviour of the sum $\sum_{|\lambda_n| < T} A_n$ for T tending to $+\infty$. By the Abel theorem which is the converse of the Tauberian theorem (41), (42) we find an asymptotic formula for the function $\sum_{n=-\infty}^{+\infty} |\lambda_n A_n|/(\lambda_n^2 + y^2)$ valid when y tends to $+\infty$. By adding this formula to (47) we get an asymptotic formula for $\sum_{n=0}^{\infty} \lambda_n A_n/(\lambda_n^2 + y^2)$. Putting $\lambda_n^2 = \mu_n$, $y^2 = r$ in this formula and applying once more the Tauberian theorem we obtain a formula from which the asymptotic behaviour of $\sum_{0<\lambda_n<T} \lambda_n A_n$ can be derived. The result is that

$$(49)\qquad \sum_{0<\lambda_n<T} \lambda_n A_n = [(H_1 \sin(\tfrac{1}{2}\pi a) + H_2 \cos(\tfrac{1}{2}\pi a))/\pi(2-a)]T^{2-a} + o(T^{2-a})$$

when T tends to $+\infty$. In a similar way we find

$$(50)\qquad \sum_{-T<\lambda_n<0} \lambda_n A_n = [(-H_1 \sin(\tfrac{1}{2}\pi a) + H_2 \cos(\tfrac{1}{2}\pi a))/\pi(2-a)]^{2-a} + o(T^{2-a})$$

for T tending to $+\infty$.

Note. We remark that the same result (49), (50) is true even if the hypothesis $A_n > 0$ is replaced by $\lambda_n A_n > 0$. To prove this one has only to interchange the rôles of (47) and (48) in the indicated deduction of (49), (50).

11. Deduction of (6) and (7). With

$$A_n = \phi_n^2(P)/|\lambda_n|, \qquad H_1 = \{\tfrac{1}{2}\,|q(P)|\}^{\frac{1}{2}}/4\pi \operatorname{sign} q(P),$$
$$H_2 = \{\tfrac{1}{2}\,|q(P)|\}^{\frac{1}{2}}/4\pi, \qquad a = \tfrac{1}{2}$$

the formula (45) takes the form (46). Corresponding to (49) we find the relation (6)

$$\sum_{0<\lambda_n<T} \phi_n^2(P) = [q_+^{\frac{3}{2}}(P)/6\pi^2]T^{3/2} + o(T^{3/2}).$$

Similarly, the relation (7) is obtained from the formula (50)

$$\sum_{-T<\lambda_n<0} \phi_n^2(P) = [q_-^{\frac{3}{2}}(P)/6\pi^2]T^{3/2} + o(T^{3/2}).$$

12. On the case when $p \equiv 0$. We indicate briefly how our results can be deduced in the case when the hypothesis min. $p(x_1, x_2, x_3) > 0$ is replaced by $p(x_1, x_2, x_3) \geqq 0$. We may as well develop the method when p vanishes identically in V.

We consider first the equation (1) with min. $p > 0$, but instead of the Green's function $G(P, Q; iy)$ we consider $G(P, Q; x + iy)$ with $x \neq 0$. In order to estimate $\gamma(P, P; x + iy)$ when $q(P) > 0$, we choose x negative. In the same way as before we get

$$M^I \geqq \gamma_1(P, P; x + iy) - D^*{}_1(\Gamma_1, \Gamma_2; \Gamma_1, \Gamma_2) \geqq M^V \tag{51}$$

but where now

$$\Gamma = \Gamma(P, Q; x + iy),$$

$$M^I = -\text{min.} \int_{V'} [\text{grad}^2 u_2 + pu_2^2 - h_1 u_2^2 + 2u_2(\Delta\Gamma_2 - p\Gamma_2 + h_1\Gamma_2 + h_2\Gamma_1)]dV,$$

$$M^V = \text{min.} \int_{V'} [\text{grad}^2 u_1 + pu_1^2 - h_1 u_1^2 + 2u_1(\Delta\Gamma_1 - p\Gamma_1 + h_1\Gamma_1 - h_2\Gamma_2)]dV.$$

In (51) we let p tend to zero, thus obtaining a similar inequality but with γ_1, Γ_1 and Γ_2 related to the equation $\Delta u + zqu = 0$ instead of (1). The obtained values of M^I and M^V are finite because of the fact that $-h_1$ is non-negative and not identically equal to zero in V'. With $p = 0$ we deduce estimates for M^I, M^V and $D^*(\Gamma, \Gamma)$ from which follows

$$\gamma_1(P, P; x + iy) = O(1) \text{ when } y \text{ tends to } +\infty.$$

In a similar way

$$\gamma_2(P, P; x + iy) = O(1) \text{ when } y \text{ tends to } + \infty$$

is deduced. The case when $q(P)$ is negative in the considered point P is treated similarly but with x positive.

This proves (6) *and* (7).

13. Asymptotic formulas for the eigenfunctions in sub-domains where $q(P) \equiv 0$. We once more consider our problems under the condition that min. $p > 0$. Let P be a point in a neighborhood of which $q(Q) \equiv 0$ and choose R so small that $q(Q) \equiv 0$ in the sphere $V' = (r_{PQ} < 0)$. Because of this choice $\Gamma(P, Q; iy)$ is independent of y. The development of **5-7** remains valid (and is simplified) and we obtain

$$\lim_{Q=P} [G(P, Q; iy) - G(P, Q)] = O(1) \text{ for } y \text{ tending to } + \infty.$$

By considerations analogous to those of **9-11** it follows *that when T tends to $+ \infty$*

$$\sum_{0 < \lambda_n < T} \phi_n^2(P) = o(T^{1-\epsilon}),$$

$$\sum_{-T < \lambda_n < 0} \phi_n^2(P) = o(T^{1-\epsilon}),$$

where ϵ is an arbitrarily small positive quantity.

Without the assumption min. $p > 0$ we cannot expect such simple asymptotic rules to be valid in sub-domains where q vanishes identically. Thus for the properly normalized eigenfunctions $\phi_n(x)$ of the one-dimensional problem

$$d^2u/dx^2 + \lambda q(x) u = 0 \text{ in } -\pi \leq x \leq \pi,$$

$$du/dx = 0 \text{ for } x = -\pi \text{ and for } x = \pi,$$

where

$$q(x) = \begin{cases} 1 \text{ in } 0 \leq x \leq \pi, \\ 0 \text{ in } -\pi \leq x < 0, \end{cases}$$

one finds

$$\lim_{T=\infty} T^{-\frac{1}{2}} \sum_{0 < \lambda_n < T} \phi_n^2(x) = \begin{cases} 1/\pi \text{ when } 0 < x < \pi \\ 2/\pi \text{ when } -\pi < x < 0. \end{cases}$$

INSTITUTE FOR ADVANCED STUDY, U. S. A AND LUND UNIVERSITY, SWEDEN.

BIBLIOGRAPHY

[1] Carleman, T., "Propriétés asymptotiques des fonctions fondamentales des membranes vibrantes," *Skandinaviska Matematikerkongressen 1934*, Stockholm, 1934.

[2] ———, "Über die asymptotische Verteilung der Eigenwerte partieller Differentialgleichungen," *Berichte über die Verhandlungen der Sächischen Akademie der Wissenschaften Leipzig*, 88, 1936.

[3] Courant, R., *Methoden der Mathematischen Physik II*, Berlin, 1937, pp. 471-544.

[4] Pleijel, Å., "Propriétés asymptotiques des fonctions et valeurs propres de certains problèmes de vibration," *Arkiv för Matematik, Astronomi och Fysik*, vol. 27 A, No. 13, 1940.

[5] ———, "Sur la distribution des valeurs propres de problèmes régis par l'équation $\Delta u + \lambda k(x, y) u = 0$," *Arkiv för Matematik, Astronomi och Fysik*, vol. 29 B, No. 7, 1943.

[6] ———, "On Hilbert-Schmidt's theorem in the theory of partial differential equations," *Fysiografiska Sällskapets i Lund Förhandlingar*, vol. 17, No. 2, 1946.

ERRATA.

C. T. Rajagopal, "On Riesz summability and summability by Dirichlet's series: addendum and corrigendum," this JOURNAL, vol. LXIX, pp. 851-852:

In condition (i*) of Corollary 3, p. 851, line 17, '$t \to +\infty$' is to be corrected to '$t \to +0$.'

C. T. Rajagopal, "Some limit theorems," this JOURNAL, vol. LXX, pp. 157-166:

Page 157, l. 4, to be read as (§ 3) instead of (**3**).

Page 158, l. 9, to be read as $\sum_{N+1}^{n}$ instead of $\sum_{n+1}^{n}$.

Page 158, l. 9, to be read as $\frac{D_N}{M_N}$ instead of $\frac{D_n}{M_n}$.

Page 158, l. 18, to be read as $[\frac{D_N}{M_N} + \cdots]$ instead of $[\frac{D_n}{M_n} + \cdots]$.

Page 159, l. 4, to be read as $\overline{\lim}$ instead of $\underline{\lim}$.

Page 159, l. 4, to be read as $\underline{\lambda}$ instead of λ.

Page 160, l. 20, to be read $\left(\frac{1}{D_\nu} - \frac{1}{D_{\nu+1}}\right)$ instead of $\left(\frac{1}{D_\nu} \quad \frac{1}{D_{\nu+1}}\right)$.

Page 160, l. 22, to be read $\left(\frac{1}{D_\nu} - \frac{1}{D_{\nu+1}}\right)$ instead of $\left(\frac{1}{D_\nu} \quad \frac{1}{D_{\nu+1}}\right)$.

Page 162, l. 21, to be read as $\underline{\lim}$ instead of lim.

Page 165, l. 27, to be read as $\sum \epsilon_n a_n$ instead of $\sum e_n a_n$.